VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÝCH SYSTÉMŮ
DEPARTMENT OF COMPUTER SYSTEMS

ANALÝZA KVALITY OVZDUŠÍ V KANCELÁŘSKÝCH A OBYTNÝCH PROSTORECH

AIR QUALITY ANALYSIS IN OFFICE AND RESIDENTIAL AREAS

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE — Bc. PETER TISOVČÍK
AUTHOR

VEDOUCÍ PRÁCE — Doc. Ing. JAN KOŘENEK, Ph.D.
SUPERVISOR

BRNO 2019

Ústav počítačových systémů (UPSY) Akademický rok 2018/2019

Zadání diplomové práce

22184

Student: **Tisovčík Peter, Bc.**
Program: Informační technologie Obor: Počítačové a vestavěné systémy
Název: **Analýza kvality ovzduší v kancelářských a obytných prostorech**
Air Quality Analysis in Office and Residential Areas
Kategorie: Vestavěné systémy
Zadání:

1. Nastudujte problematiku měření kvality ovzduší. Zaměřte se na koncentraci oxidu uhličitého.
2. Seznamte se s principy fungování systému BeeeOn.
3. Pro vybrané místnosti nadefinujte parametry ovlivňující koncentraci oxidu uhličitého (přítomnost osob, rostlin, atd.).
4. Vyberte vhodnou sadu senzorů pro měření koncentrace oxidu uhličitého, ve vhodných místnostech nasaďte senzory a proveďte dlouhodobé měření.
5. Navrhněte systém, který bude na základě měřených dat detekovat události spojené se změnou oxidu uhličitého a navrhovat vhodný způsob regulace výměny vzduchu s cílem zlepšit kvalitu ovzduší.
6. Navržený systém implementujte a ověřte nad reálnými daty.
7. V závěru diskutujte dosažené výsledky.

Literatura:

- Dle pokynů vedoucího.

Při obhajobě semestrální části projektu je požadováno:

- Splnění bodů 1 až 3 zadání.

Podrobné závazné pokyny pro vypracování práce viz http://www.fit.vutbr.cz/info/szz/

Vedoucí práce: **Kořenek Jan, Ing., Ph.D.**
Vedoucí ústavu: Sekanina Lukáš, prof. Ing., Ph.D.
Datum zadání: 1. listopadu 2018
Datum odevzdání: 22. května 2019
Datum schválení: 26. října 2018

Abstrakt

Cieľom práce bolo naštudovať problematiku merania kvality vnútorného ovzdušia, so zameraním na koncentráciu oxidu uhličitého. V rámci teoretickej časti práce bola popísaná problematika získavania znalostí z databáz, boli tiež predstavené základné klasifikačné metódy a prístupy pre redukciu dimenzionality. Ďalej boli naštudované princípy fungovania vyvíjaného systému v rámci projektu IoTCloud a dostupné možnosti pre meranie potrebných veličín. V praktickej časti práce boli vybrané vhodné senzory pre zvolené miestnosti a bolo vykonané dlhodobé meranie. Namerané dáta boli použité pre vytvorenie systému na detekciu otvorenia okna i pre návrh vhodného spôsobu regulácie výmeny vzduchu v miestnosti. Cieľom regulácie bolo zlepšiť kvalitu ovzdušia za pomoci prirodzeného vetrania.

Abstract

The goal of the thesis was to study the indoor air quality measurement focusing on the concentration of carbon dioxide. Within the theoretical part, data mining including basic classification methods and approaches to dimensionality reduction was introduced. In addition, the principles of the developed system within IoTCloud project and available possibilities for measurement of necessary quantities were studied. In the practical part, the suitable sensors for given rooms were selected and long-term measurement was performed. Measured data was used to create the system for window opening detection and for the design of appropriate way of air change regulation in a room. The aim of regulation was to improve air quality using natural ventilation.

Kľúčové slová

BeeeOn, IoTCloud, získavanie znalostí, redukcia dimenzionality, oxid uhličitý, CO_2, kvalita ovzdušia, klasifikácia, predikcia, regulácia výmeny vzduchu.

Keywords

BeeeOn, IoTCloud, data mining, dimensionality reduction, carbon dioxide, CO_2, air quality, classification, prediction, air change regulation.

Citácia

TISOVČÍK, Peter. *Analýza kvality ovzduší v kancelářských a obytných prostorech.* Brno, 2019. Diplomová práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Doc. Ing. Jan Kořenek, Ph.D.

Analýza kvality ovzduší v kancelářských a obytných prostorech

Prehlásenie

Prehlasujem, že som túto diplomovú prácu vypracoval samostatne pod vedením pána Doc. Ing. Jana Kořenka, Ph.D. Uviedol som všetky literárne pramene a publikácie, z ktorých som čerpal.

.....
Peter Tisovčík
22. mája 2019

Poďakovanie

Rád by som poďakoval vedúcemu diplomovej práce Doc. Ing. Janovi Kořenkovi, Ph.D. za poskytnuté odborné konzultácie, či už pri písaní textovej časti práce alebo návrhu. Práca bola vytvorená ako súčasť projektu IoTCloud, ktorý bol spolufinancovaný Technologickou agentúrou Českej republiky.

Obsah

1 Úvod **5**

2 Kvalita ovzdušia **7**
2.1 Vonkajšie ovzdušie . . . 7
2.2 Vnútorné ovzdušie . . . 8
2.3 Syndróm nezdravých budov . . . 9
2.4 Oxid uhličitý . . . 10
2.4.1 Pettenkoferovo kritérium . . . 10
2.4.2 Vznik oxidu uhličitého a oxidu uhoľnatého . . . 10
2.4.3 Využitie oxidu uhličitého . . . 11
2.4.4 Vplyv oxidu uhličitého na zdravie ľudí . . . 12
2.5 Intenzita vetrania a výmeny vzduchu . . . 13
2.6 Hodnotenie vnútorného prostredia budov . . . 15

3 Získavanie znalostí z databáz **16**
3.1 Proces získavania znalostí . . . 17
3.2 Typy úloh pre dolovanie dát . . . 18
3.3 Klasifikačné a predikčné metódy . . . 19
3.3.1 Rozhodovací strom . . . 19
3.3.2 Náhodný les . . . 21
3.3.3 Bayesovská klasifikácia . . . 22
3.3.4 Algoritmus podporných vektorov . . . 23
3.3.5 Umelé neurónové siete . . . 24
3.4 Nástroje na dolovanie dát . . . 25

4 Redukcia dimenzionality **26**
4.1 Extrakcia atribútov . . . 27
4.2 Výber atribútov . . . 27
4.2.1 Wrapper metódy . . . 27
4.2.2 Filtračné metódy . . . 28
4.2.3 Embedded metódy . . . 31

5 Merania oxidu uhličitého **32**
5.1 Prístroje pre meranie koncentrácie oxidu uhličitého . . . 32
5.2 Snímače zmesi plynov . . . 35
5.3 BeeeOn systém . . . 35
5.4 Výber senzorov . . . 36

6 Zber a spracovanie dát **38**
6.1 Testovacie prostredia 38
6.1.1 Hlavné testovacie prostredie 38
6.1.2 Ostatné testovacie prostredia 39
6.2 Spracovanie nameraných hodnôt 40
6.3 Tvorba a výber atribútov 41

7 Detekcia otvorenia okna **43**
7.1 Detekcia na základe koncentrácie oxidu uhličitého 45
7.2 Detekcia na základe koncentrácie oxidu uhličitého, vnútornej teploty a vlhkosti 48
7.3 Detekcia na základe koncentrácie oxidu uhličitého, vnútornej a vonkajšej teploty a vlhkosti 51

8 Regulácia vzduchu v miestnosti **54**

9 Záver **57**

Literatúra **59**

A Obsah priloženého pamäťového média **63**

B Úspešnosti detektorov pri vyhodnotení obecnosti **64**

C Úspešnosť klasifikátorov pri určení intenzity výmeny vzduchu **71**

Zoznam tabuliek

2.1 Zloženie atmosferického vzduchu [31] 7
2.2 Druhy problémov v 203 budovách s ohľadom na kvalitu vnútorného ovzdušia vykonané NIOSH [7] 9
2.3 Množstvo vyprodukovaného oxidu uhličitého v závislosti od činnosti človeka a zdroja dát [12] 11
2.4 Vplyv koncentrácie CO_2 na ľudský organizmus [36] 12
2.5 Základná klasifikácia kvality vnútorného ovzdušia podľa normy ČSN EN 16798-3 15

5.1 Parametre senzorov od firiem Protronix a IQ Home [13, 25] 37

7.1 Dosiahnuté úspešnosti klasifikátorov pri počiatočnom vyhodnotení v hlavnom testovacom prostredí (uvažovaná len koncentrácia CO_2) 45
7.2 Dosiahnuté úspešnosti klasifikátorov po 58. iterácii v hlavnom testovacom prostredí (uvažovaná len koncentrácia CO_2) 46
7.3 Úspešnosť náhodného lesa pri obecnom otestovaní v hlavnom testovacom prostredí (uvažovaná len koncentrácia CO_2) 46
7.4 Úspešnosť náhodného lesa pri obecnom otestovaní v kancelárskej miestnosti (uvažovaná len koncentrácia CO_2) 47
7.5 Úspešnosť neurónovej siete pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná len koncentrácia CO_2) 47
7.6 Úspešnosť neurónovej siete pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná len koncentrácia CO_2) 47
7.7 Dosiahnuté úspešnosti klasifikátorov pri počiatočnom vyhodnotení v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť) 48
7.8 Dosiahnuté úspešnosti klasifikátorov po 58. iterácii v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť) 49
7.9 Úspešnosť náhodného lesa pri obecnom otestovaní v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť) 49
7.10 Úspešnosť rozhodovacieho stromu pri obecnom otestovaní v kancelárskom prostredí (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť) 50
7.11 Úspešnosť náhodného lesa pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť) 50
7.12 Úspešnosť neurónovej siete pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť) 50
7.13 Dosiahnuté úspešnosti klasifikátorov pri počiatočnom vyhodnotení v hlavnom testovacom prostredí (uvažovaná koncentrácia CO, vnútorná a vonkajšia teplota a vlhkosť) 51

7.14 Dosiahnuté úspešnosti klasifikátorov po 60. iterácii v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť) 52
7.15 Úspešnosť náhodného lesa pri obecnom otestovaní v hlavnej testovacej miestnosti (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť) 52
7.16 Úspešnosť neurónovej siete pri obecnom otestovaní v kancelárskej miestnosti (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť) . . . 53
7.17 Úspešnosť náhodného lesa pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť) . . . 53
7.18 Úspešnosť neurónovej siete pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť) . . . 53

B.1 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v hlavnom testovacom prostredí (uvažovaná len koncentrácia CO_2) 64
B.2 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v kancelárskej miestnosti (uvažovaná len koncentrácia CO_2) 65
B.3 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná len koncentrácia CO_2) 65
B.4 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná len koncentrácia CO_2) 66
B.5 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť) 66
B.6 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v kancelárskej miestnosti (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť) 67
B.7 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť) 67
B.8 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť) 68
B.9 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť) 68
B.10 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v kancelárskej miestnosti (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť) 69
B.11 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť) 69
B.12 Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť) 70

C.1 Úspešnosť klasifikátorov pri určení intenzity výmeny vzduchu v závislosti na počte košov a spôsobu vyhodnotenia úspešnosti 72

Kapitola 1

Úvod

Dôsledkom modernej doby ľudia trávia väčšinu času v kanceláriách, školách alebo iných obytných priestoroch. V týchto priestoroch dýchame vnútorný vzduch, ktorý je ovplyvnený kvalitou vonkajšieho vzduchu. Dôvodom tohto ovplyvňovania je vetranie, ktoré privádza vonkajší čerstvý vzduch do budov a odvádza vzduch obsahujúci znečisťujúce látky von z budov. Vnímanie kvality vnútorného ovzdušia je subjektívne hodnotenie, ktoré sa môže skladať z rôznych faktorov ako kvalita vonkajšieho vzduchu, objem vzduchu pripadajúceho na jednu osobu v miestnosti, alebo aj ako množstvo znečisťujúcich látok v ovzduší. Zdrojom znečisťujúcich látok vo vnútorných priestoroch môžu byť obyvatelia, zvieratá, rastliny ale aj samotný materiál, z ktorého je budova postavená.

Mnoho budov má nútenú výmenu vzduchu podľa dopredu daného časového harmonogramu a počtu očakávaných ľudí vo vybraných priestoroch. V prípade zvýšenia počtu ľudí nad neočakávanú hranicu môže dôjsť k nedostatočnej výmene vzduchu a postupnému zhoršovaniu kvality vnútorného prostredia. Opačným prípadom je zbytočné privádzanie veľkého objemu čerstvého vzduchu do priestorov, kde sa nachádza výrazne menej osôb, ako bolo plánované.

Vďaka snahe znižovať energetickú náročnosť budov, dochádza k používaniu nových materiálov, ktoré eliminujú infiltráciu vonkajšieho vzduchu. Infiltráciou sa rozumie prirodzené vetranie pomocou netesností vonkajšej obálky budovy (napr. netesné škáry okenných a dverných otvorov). Moderné technológie a materiály obmedzujú túto infiltráciu tak výrazne, že bez núteného alebo prirodzeného vetrania dochádza k rýchlemu nárastu nežiadúcich látok vo vnútornom prostredí, čo môže spôsobiť zdravotné problémy.

Celková kvalita vnútorného ovzdušia ovplyvňuje naše fyzické zdravie a duševnú pohodu. Zdravotné problémy môže spôsobiť zlá kvalita vnútorného ovzdušia, nevyhovujúca teplota alebo vlhkosť. Nevyhovujúca teplota a vlhkosť spolu s nedostatočným vetraním môžu spôsobiť rast plesní a ďalších alergénov.

Podľa vyhlášky č. 20/2012 Sbírky zákonů Českej republiky o technických požiadavkách na stavby je ukazovateľom kvality vnútorného prostredia koncentrácia oxidu uhličitého, ktorá by nemala presiahnuť hranicu 1 500 ppm. Po prekročení tejto hodnoty môže dochádzať k únave, bolestiam hlavy, alebo hrdla. Dlhodobé pôsobenie vysokej koncentrácie oxidu uhličitého na ľudský organizmus môže viesť k vzniku astmy a ďalších ochorení dýchacích ciest. Pričom vysoká koncentrácia oxidu uhličitého môže mať na ľudí rozdielne účinky.

Koncentrácia oxidu uhličitého vo vnútornom prostredí závisí najmä od množstva a kvality privádzaného čerstvého vzduchu a od počtu zdrojov nachádzajúcich sa v danom prostredí. Množstvo oxidu uhličitého, ktoré osoby vyprodukujú, závisí od ich činnosti a hmotnosti. Čím je vykonávaná fyzicky náročnejšia práca, tým sa u ľudí spaľuje viacej glukózy,

pričom sa vytvára väčšie množstvo oxidu uhličitého a jeho koncentrácia rastie rýchlejšie. Koncentráciu ovplyvňujú aj rastliny, ktoré fotosyntézou premieňajú oxid uhličitý na kyslík.

V dnešnej modernej dobe sú vetracie systémy uspôsobené na určitý počet osôb v miestnosti, prípadne na časový harmonogram. Takéto uspôsobenie však nie je efektívne v prípade nečakaného zvýšenia počtu osôb v miestnosti, alebo v prípade, kedy sa z nejakých príčin (dovolenka, chrípka) na pracovisku vyskytuje podstatne menej osôb ako obyčajne. Niektoré budovy z rozličných dôvodov nemôžu mať ventilačné systémy, a preto musí byť vetranie vykonávané prirodzene (ručné otvorenie okna). Informáciu o tom, ako dlho a kedy sa má vetrať, je len subjektívny pohľad danej osoby. V budovách môžu byť umiestnené meracie prístroje (senzory), ktoré umožňujú merať, prípadne aj signalizovať, koncentráciu oxidu uhličitého. Neumožňujú však predpovedať, za aký čas je treba miestnosť vyvetrať, aby koncentrácie oxidu uhličitého neprekročila stanovené hranice, alebo ako dlho by toto vetranie malo prebiehať, aby bolo vnútorné prostredie dostatočne vyvetrané. Ďalej je možné použiť senzory oxidu uhličitého aj ako bezpečnostný prvok pre detekciu otvorenia alebo zatvorenia okien a dverí.

Cieľom práce bolo vybrať vhodné senzory na meranie oxidu uhličitého, teploty a vlhkosti vo vnútri budov a na základe vykonaných meraní stanoviť závislosti koncentrácie oxidu uhličitého na daných veličinách. Ďalším cieľom bolo na základe stanovených závislostí vytvoriť alebo vybrať vhodné parametre pre modely na detekciu otvorenia okna. Pre odhad doby, ako dlho má vetranie v miestnosti prebiehať, aby bola dosiahnutá požadovaná koncentrácia oxidu uhličitého, bolo tiež potreba vybrať alebo vytvoriť vhodné parametre.

Text práce je rozčlenený do niekoľkých kapitol. V druhej kapitole sú uvedené informácie o kvalite ovzdušia, parametre pre hodnotenie kvality vnútorného ovzdušia, ktorá je ovplyvnená hlavne koncentráciou oxidu uhličitého. Okrem obecných informácii o kvalite vzduchu sú uvedené informácie o možnostiach vzniku oxidu uhličitého, o jeho využití, vplyvu na ľudský organizmus a možnostiach zníženia jeho koncentrácie vo vnútorných priestoroch. Koniec kapitoly je venovaný hodnoteniu vnútorného prostredia budov a prípustným limitným hodnotám, ktorých môže oxid uhličitý v budovách dosahovať. Tretia kapitola popisuje oblasť získavania znalostí z databáze, vrátane informácii o procese pre získanie výsledkov dolovania z dát. Koniec kapitoly je zameraný na riešenie dolovacích úloh pomocou najpoužívanejších metód, ako napríklad rozhodovacieho stromu a neurónovej siete. V ďalšej kapitole je zhrnutá problematika redukcie dimenzionality a základné metódy pre zníženie dimenzionality. Piata kapitola je venovaná princípom merania koncentrácie oxidu uhličitého rôznymi metódami, ktoré majú svoje výhody a nevýhody prameniace z použitej technológie. Ďalej je popísaný systém vyvíjaný v rámci projektu IoTCloud, ktorý spracováva namerané dáta zo senzorov. Použité senzory sú popísané v závere kapitoly. V ďalšej kapitole, ktorá sa zaoberá vyhodnotením nameraných dát, sú popísané testovacie prostredia, v ktorých boli senzory nasadené, a princíp spracovania nameraných hodnôt. Ďalej je popísaná tvorba a výber atribútov slúžiacich pre následné vytvorenie detektora otvorenia okna a tiež pre návrh regulácie vzduchu v miestnosti. V záverečnej kapitole sú zhrnuté dosiahnuté výsledky a sú navrhnuté ďalšie možné vylepšenia.

Kapitola 2

Kvalita ovzdušia

Vo všeobecnosti závisí kvalita vnútorného ovzdušia na kvalite vonkajšieho ovzdušia, pretože vetraním privádzame do budov vonkajší vzduch. Na jednej strane sa vetraním odvádzajú škodliviny vzniknuté v budove a na druhej strane sa privádzajú škodliviny z vonkajšieho ovzdušia do budovy.

Mieru kvality ovzdušia upravuje zákon a ďalej ju špecifikujú vyhlášky, ktoré sú špecifické pre každú krajinu. Pre Českú republiku je to vyhláška č. 20/2012 Sbírka zákonů. Táto vyhláška určuje oxid uhličitý ako ukazovateľ kvality ovzdušia [28]. Okrem zákonov a vyhlášok sú v normách dostupné aj doplňujúce predpisy. Normy, zaoberajúce sa kvalitou ovzdušia, ponúkajú limitné a odporúčané hodnoty napríklad pre koncentrácie oxidu uhličitého alebo dávku čerstvého vzduchu na osobu.

2.1 Vonkajšie ovzdušie

Kvalita vonkajšieho ovzdušia je určená množstvom znečisťujúcich látok, ktoré obsahuje. Znečistenie spôsobujú hlavne rôzne druhy spaľovacích procesov, ktoré je možné pozorovať v priemysle, pri výrobe energií a v doprave. Medzi látky, ktoré najviac znečisťujú vonkajšie ovzdušie v Českej republike, patria tuhé látky (oxid siričitý, oxidy dusíka, oxid uhoľnatý, polyaromatické uhľovodíky, prchavé organické látky a amoniak) [10].

Atmosferický vzduch sa skladá z niekoľkých významných plynných zložiek. Hlavnou plynnou zložkou vzduchu je dusík (78,09 %), kyslík (20,95 %), argón (0,93 %) a oxid uhličitý (0,03 %). Okrem týchto plynov sa môžu vo vzduchu nachádzať aj ďalšie plyny, ale ich obsah je zanedbateľný [31]. Obsah plynu môže byť vyjadrený objemovo alebo hmotnostne, ako je zobrazené v tabuľke 2.1.

Plynná zložka	**Obsah podľa objemu [%]**	**Obsah podľa hmotnosti [%]**
dusík	78,09	75,50
kyslík	20,95	23,17
argón	0,93	1,286
oxid uhličitý	0,03	0,043

Tabuľka 2.1: Zloženie atmosferického vzduchu [31]

Ovzdušie obsahujúce znečisťujúce látky môže mať významné zdravotné dopady v podobe výskytu rôznych chorôb a problémov, prípadne zhoršenie existujúcich príznakov, ktoré sú spojené s dýchacími cestami [18].

2.2 Vnútorné ovzdušie

V prostredí, ktoré sa nachádza vo vnútri budov, trávime najviac času, a to približne 90 % [17]. Toto prostredie má významný vplyv na naše fyzické zdravie. Problém nastáva s kvalitou vnútorného ovzdušia, ktorá je daná kombináciou vnútorných zdrojov a prenikaním škodlivín z vonkajšieho ovzdušia. Dominantný význam majú faktory prostredia, ktoré v budovách zapríčiňujú rast pliesní, alergénov a predstavujú zdravotné riziko pre osoby nachádzajúce sa v týchto priestoroch [17]. Kvalitu vnútorného prostredia ovplyvňujú nasledujúce faktory [15]:

- fyzikálne – prúdenie vzduchu, hluk, teplota a vlhkosť,
- chemické – oxid uhličitý, uhoľnatý a dusičitý, anorganické a organické škodliviny, prchavé organické látky, azbest a fajčenie,
- biologické – baktérie, vírusy, plesne, roztoče, prach, domáce zvieratá a ľudia.

Hodnotenie kvality vnútorného ovzdušia je subjektívne a závisí na [30]:

- kvalite vonkajšieho ovzdušia,
- objemu vzduchu pripadajúceho na jednu osobu v miestnosti,
- výmene vzduchu,
- množstve znečisťujúcich látok v ovzduší, kde týmto zdrojom môžu byť:
  - ľudia, zvieratá, rastliny a ich metabolizmus,
  - aktivity ľudí a zvierat,
  - stavebný materiál,
  - upratovanie, čistenie a údržba budov.

Aby bola kvalita vnútorného ovzdušia čo najlepšia, je potrebné dostatočné vetranie v pravidelných časových intervaloch, alebo trvalé vetranie s neustálou výmenou vzduchu. Typy vetrania sa delia na dve základné kategórie – prirodzené a mechanické (nútené) vetranie. Dominantou prirodzeného vetrania je využitie prenikania vzduchu otvorenými oknami alebo netesnosťami v plášti budovy (infiltrácia). Mechanické vetranie využíva na rozdiel od prirodzeného vetrania ventilátory, ktoré zabezpečujú odvod a prívod vzduchu. V prípade použitia mechanického vetrania len na odvod alebo len na prívod vzduchu je nutné použiť aj prirodzené vetranie [9].

2.3 Syndróm nezdravých budov

Syndróm nezdravých budov (SBS – Sick Building Syndrom) je pojem používaný pre nešpecifikované symptómy súvisiace s pobytom v budovách. Ľudia vo vnútri budov sa sťažujú na symptómy ako podráždenie očí, nosu, krku, bolesti hlavy, krku, únavu, alergiu (prípadne prechod do astmy), suchú pokožku, horúčky alebo kašeľ [7].

Jednotliví ľudia sa môžu sťažovať na rozdielne symptómy. Ak je počet sťažujúcich sa ľudí väčší než 20 % po dobu aspoň dvoch týždňov, je budova označená ako chorá [7].

Príčiny vzniku SBS je možné nájsť v nových konštrukciách budov, zatepľovaní budov alebo v používaní plastových okien. Tieto nové konštrukcie a materiály šetria energiu a umožňujú menšiu infiltráciu vonkajšieho vzduchu. Okrem samotných budov a materiálov, z ktorých sú zhotovené, prispievajú k znečisťovaniu vnútorného prostredia technológie (počítače, tlačiarne), nábytok, koberce, či samotné vetracie systémy, ktoré je potrebné pravidelne kontrolovať a čistiť [7].

Pôvod znečistenia	**Počet výskytov**	**Výskyt [%]**
nedostatočné vetranie	98	48
kontaminácia (vnútorná)	36	18
kontaminácia (vonkajšia)	21	10
neznáme	19	9
vlhkosť	9	4
kontaminácia (konštrukcia budovy)	7	3
mikrobiologická kontaminácia	6	3
cigaretový dym	4	2
hluk, osvetlenie	2	1
svrab	1	2

Tabuľka 2.2: Druhy problémov v 203 budovách s ohľadom na kvalitu vnútorného ovzdušia vykonané NIOSH [7]

Výskumníci z NIOSH[1] vykonali výskum v USA roku 1983, v rámci ktorého bolo otestovaných 203 budov. Z výskumu vyplýva, že hlavnou príčinou SBS je nedostatočné vetranie, ktoré tvori skoro polovicu (48 %) celkového znečistenia budov [7], ako je možné vidieť v tabuľke 2.2.

[1] National Institute for Occupational Safety and Health

2.4 Oxid uhličitý

Oxid uhličitý, tiež nazývaný kysličník uhličitý, je stály, bezfarebný, nehorľavý, atmosferický plyn bez chuti a zápachu. Koncentrácia oxidu uhličitého v málo osídlených miestach je 350 ppm, v mestách 400 ppm a okolo 450 ppm v centre miest, pri dopravných tepnách alebo v priemyselných zónach [12]. Oxid uhličitý sa skladá z jedného atómu uhlíka a dvoch atómov kyslíka, často sa označuje svojou rovnicou CO_2 a je približne 1,5-krát ťažší než vzduch [14].

Koncentrácia oxidu uhličitého sa najčastejšie udáva v jednotkách *ppm* (*parts per milion*), ktoré vyjadrujú počet častíc látky na milión iných častíc. Niekedy je však výhodnejšie vyjadriť koncentráciu v hmotnostných jednotkách ($mg{\cdot}m^{-3}$), ktoré udávajú hmotnosť plynu v určitom objeme. Hodnota koncentrácie CO_2 v rôznych jednotkách je nasledujúca [8]:

$$1000\,ppm \;=\; 0,1\,\% \;=\; 1800\,mg \cdot m^{-3}$$

2.4.1 Pettenkoferovo kritérium

Max Joseph von Pettenkofer preukázal, že hlavnými produktami látkovej výmeny človeka je oxid uhličitý a vodná para. Meraním množstva CO_2 vo vydýchanom vzduchu zistil, že produkcia CO_2 závisí na fyzickej aktivite. Osoba v bdelom stave produkuje približne $16\,l \cdot h^{-1}$ CO_2. Ďalej zistil, že koncentrácia CO_2 udáva kvalitu vnútorného prostredia a stanovil maximálne prípustné množstvo CO_2 na 1000 ppm. Na základe zistení definoval dávku čerstvého vzduchu pre jednu dospelú osobu na $25\,m^3 \cdot h^{-1}$ [21].

2.4.2 Vznik oxidu uhličitého a oxidu uhoľnatého

Oxid uhličitý vzniká hlavne pri spaľovacích procesoch reakciou uhlíka a kyslíka, alebo ako produkt metabolizmu živých organizmov. Prirodzeným procesom vo voľnej prírode vzniká únikom zo sopky, hnitím, kvasením, dýchaním alebo dôsledkom ľudskej činnosti (antropogénne CO_2) [4]. Antropogénne CO_2 sa vytvára predovšetkým počas spaľovania fosílnych a organických palív. Spaľovanie palív je dej, pri ktorom sa zlučujú prvky, ktoré sa nachádzajú v palive, s kyslíkom, ktorý je súčasťou atmosferického vzduchu [22]. Počas dokonalého spaľovacieho procesu sa uhlík v palive zlúči s kyslíkom vo vzduchu a vznikne oxid uhličitý podľa nasledujúcej rovnice

$$C + O_2 \longrightarrow CO_2. \tag{2.1}$$

Oxid uhoľnatý (CO) vzniká pri nedokonalom spaľovaní fosílnych palív, pri požiari v uzavretých priestoroch alebo sa môže nachádzať vo výfukových plynoch. Pre človeka je CO toxický a vo vyšších koncentráciách môže spôsobiť až smrť. Dôvodom je, že CO sa viaže v krvi na hemoglobín 250–300-krát pevnejšie než kyslík, a tak znemožňuje hemoglobínu v transporte kyslíka z pľúc do tkanív [22]. CO vzniká rovnicou

$$2\,C + O_2 \longrightarrow 2\,CO. \tag{2.2}$$

Hlavným zdrojom CO_2 vo vnútri budov je respirácia, čo je proces získavania energie rozložením cukrov na bunkovej úrovni. Proces respirácie môžeme nazvať dýchaním, ak sa jedná o ľudí, prípadne fotosyntézou, ak sa jedná o rastliny.

Reakciou glukózy a kyslíka pri dýchaní vznikajú produkty – oxid uhličitý, voda a ATP (adenozín trifosfát – energia) [2], ako ukazuje rovnica

$$C_6H_{12}O_6 + 6\ O_2 \longrightarrow 6\ H_2O + 6\ CO_2. \tag{2.3}$$

Rýchlosť respirácie záleží od typu činnosti. Čím je činnosť fyzicky náročnejšia, tým sa frekvencia dýchania zvyšuje a človek tak prijíma a vylučuje väčšie množstvo vzduchu. Produkcia CO_2 závisí nielen od fyzickej činnosti, ale aj od výšky osoby, jej hmotnosti a od psychickej pohody.

Objem oxidu uhličitého, ktorý môže osoba vyprodukovať, rastie so zvyšujúcou sa fyzickou aktivitou a je závislý tiež od metódy a spôsobu merania. V tabuľke 2.3 je zobrazená produkcia CO_2 získaná z rôznych zdrojov pri rozličných činnostiach [12].

Pri pokojnom dýchaní, čo je približne 15 vdychov za minútu, potrebuje človek asi 7,5 litra čerstvého vzduchu za minútu. Vydychovaný vzduch človekom obsahuje približne 6,27 % vodnej pary, 15,33 % kyslíka, 4,4 % oxidu uhličitého, 75,33 % dusíka a ostatných plynov. Množstvo vyprodukovaného CO_2 jednou osobou je okolo 10–75 litrov za hodinu v závislosti na vykonávanej aktivite. Spotreba kyslíka ľudským organizmom je pomerne malá v porovnaní s CO_2, a to približne 20–25 litrov za hodinu [1].

Zdroj dát	**Produkcia CO_2** $[l \cdot h^{-1}]$	**Typ činnosti**
Recknagel, Sprenger, Schramek (1999)	20	ľahká aktivita v sede
VDI 4300 Bl. 9:2003	15–20	činnosť v sede
	20–40	ľahká práca
	40–70	stredne ťažká práca
	70–110	ťažká práca
ASHRAE (1989)	18	kancelárska práca
Max Joseph von Pettenkofer	16	bdelý stav

Tabuľka 2.3: Množstvo vyprodukovaného oxidu uhličitého v závislosti od činnosti človeka a zdroja dát [12]

2.4.3 Využitie oxidu uhličitého

V rastlinách okrem procesu respirácie, pri ktorom vzniká CO_2, prebieha aj proces fotosyntézy, čo je opačný dej ako respirácia. Na vstupe fotosyntézy je voda, oxid uhličitý, chlorofyl a energia (svetlo), jej produktom je kyslík a glukóza [2], ktoré vzniknú nasledujúcou reakciou

$$6\ H_2O + 6\ CO_2 \longrightarrow C_6H_{12}O_6 + 6\ O_2. \tag{2.4}$$

Okrem procesu fotosyntézy je možné CO_2 využiť v bežnom živote, a to pri [6]:

- extrakcií chmeľu pre použitie v pivovarníctve,
- odstraňovaní kofeínu z kávy a čaju,
- výrobe sýtených nápojov,
- ochrane výrobkov pred pôsobením kyslíka ako ochrannú atmosféru,
- napĺňaní sprejov, kde nahradzuje plyny, ktoré poškodzujú ozónovú vrstvu,
- napĺňaní hasiacich prístrojov suchým ľadom.

2.4.4 Vplyv oxidu uhličitého na zdravie ľudí

Obsah oxidu uhličitého v atmosférickom vzduchu je nízky (0,03 %) a nepredstavuje zdravotné problémy pri pobytu vo vonkajšom prostredí. Problémom sú uzavreté miestnosti, kde koncentrácia CO_2 môže rýchlo narásť, čo vedie k zvýšenej únave a k zníženej produktivite osôb nachádzajúcich sa v daných priestoroch.

Vplyv CO_2 na ľudský organizmus je tiež daný dobou, po ktorú sú osoby vystavené vysokým koncentráciám. V závislosti od dĺžky a frekvencie pobytu v prostredí s vysokou koncentráciou sa môžu vyskytnúť bolesti hlavy, respiračné problémy alebo ochorenia dýchacích ciest. Spomenuté zdravotné komplikácie môžeme pozorovať v priestoroch, ktoré sú nedostatočne vetrané, alebo v priestoroch s väčším počtom ľudí (divadlá, auly, školy, kancelárie) [36].

Účinky CO_2 na ľudský organizmus záležia od jeho koncentrácie vo vzduchu v daných priestoroch (tabuľka 2.4) a môžu mať rozdielne dopady na jednotlivé osoby. Doporučená maximálna úroveň CO_2 je stanovená v rozsahu 1 200–1 500 ppm [28]. Za bezpečnú hranicu koncentrácie CO_2 sa považuje hodnota do 5 000 ppm, ktorá nespôsobuje osobám vážne zdravotné problémy.

Koncentrácia CO_2	**Vplyv CO_2 na ľudský organizmus**
350–450 ppm	úroveň vo vonkajšom prostredí
do 1 000 ppm	doporučená úroveň CO_2 vo vnútorných priestoroch
1 200–1 500 ppm	doporučená maximálna úroveň CO_2 vo vnútorných priestoroch
1 000–2 000 ppm	príznaky únavy a znižovanie koncentrácie
2 000–5 000 ppm	bolesti hlavy
$\leq$ 5 000 ppm	maximálna bezpečná koncentrácia bez zdravotných rizík
$>$ 5 000 ppm	nevoľnosť a zvýšený tep
$>$ 15 000 ppm	dýchacie problémy
$>$ 40 000 ppm	možná strata vedomia

Tabuľka 2.4: Vplyv koncentrácie CO_2 na ľudský organizmus [36]

2.5 Intenzita vetrania a výmeny vzduchu

Aby bolo možné z kvantitatívneho hľadiska vyjadriť, aký objem vzduchu sa dostáva do miestnosti a aký podiel miestnosti sa vyvetral, boli definované dve veličiny – intenzita vetrania a intenzita výmeny vzduchu.

Intenzita vetrania je veličina, ktorú je možné vypočítať ako podiel prietoku čerstvého vonkajšieho vzduchu, ktorý prúdi do miestnosti, ku objemu danej miestnosti [37]. Veličina udáva, koľkokrát za hodinu je znečistený vzduch v miestnosti vymenený za čerstvý. Pre výpočet intenzity vetrania je možné použiť rovnicu

$$n = \frac{Q}{V}, \tag{2.5}$$

kde jednotlivé parametre n, Q a V znamenajú:

n – intenzita vetrania $[h^{-1}]$ (niekedy označovaná ako λ),

Q – intenzita výmeny vzduchu $[m^3 \cdot h^{-1}]$,

V – objem miestnosti $[m^3]$.

Druhou spomínanou veličinou je intenzita výmeny vzduchu, ktorá je definovaná ako podiel prietoku vzduchu vstupujúceho do miestnosti ku objemu danej miestnosti. Veličina vyjadruje, koľkokrát dôjde za hodinu k výmene vzduchu v danej miestnosti. Pre výpočet intenzity výmeny vzduchu je možné použiť rovnicu hmotnostnej bilancie [19], ktorá je formulovaná ako

$$\frac{V \cdot dC_i(t)}{dt} = -(C_i - C_a) \cdot Q + E, \tag{2.6}$$

kde jednotlivé parametre V, C_i, C_a, Q, E a t znamenajú:

V – objem miestnosti $[m^3]$,

C_i – koncentrácia plynu vo vnútornom vzduchu $[kg \cdot m^{-3}]$,

C_a – koncentrácia plynu v okolnom vzduchu $[kg \cdot m^{-3}]$,

Q – intenzita výmeny vzduchu $[m^3 \cdot h^{-1}]$,

E – množstvo vyprodukovaných emisií $[kg \cdot m^{-3}]$,

t – čas [h].

Po podelení oboch strán rovnice (2.6) objemom miestnosti je získaná obyčajná diferenciálna rovnica, ktorá popisuje zmenu koncentrácie plynu v miestnosti za určitý čas [19]. Pokiaľ je podiel $\frac{Q}{V}$ nahradený λ, rovnicu je možné zapísať v tvare

$$\frac{dC_i(t)}{dt} = -(C_i - C_a) \cdot \lambda + \frac{E}{V}. \tag{2.7}$$

Rovnica (2.7) vyjadruje, že pri konštantnom homogénnom zmiešaní je zmena koncentrácie plynu úmerná rozdielu koncentrácií medzi vnútorným a vonkajším vzduchom $(C_i - C_a)$ v čase t, intenzite vetrania λ a množstve vyprodukovaných emisií E [19].

Integrácia diferenciálnej rovnice vedie na exponenciálnu funkciu s počiatočnou podmienkou v čase 0 ($C(t = 0) = C_0$) a intenzitou vetrania λ v exponente. Výsledkom úprav je rovnica

$$C_i(t) = (C_0 - C_a) \cdot e^{-\lambda \cdot t} + C_a + \frac{E}{\lambda \cdot V}(1 - e^{-\lambda \cdot t}). \tag{2.8}$$

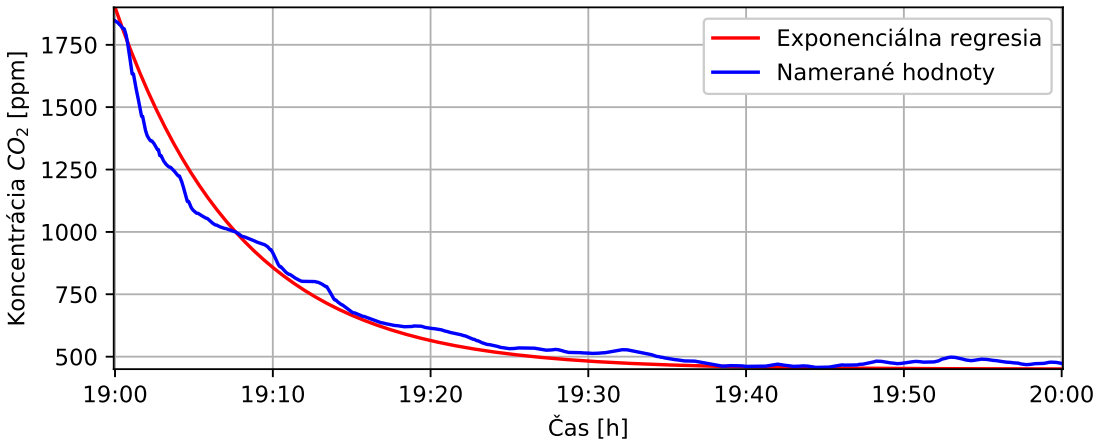
V prípade, že v miestnosti nie je žiaden zdroj emisii ($E = 0$) a počiatočná koncentrácia plynu v miestnosti C_0 je nenulová, potom výraz (2.9) popisuje odstraňovanie plynu z miestnosti

$$C(t) = (C_0 - C_a) \cdot e^{-\lambda \cdot t} + C_a, \tag{2.9}$$

kde parameter $C(t)$ znamená:

$C(t)$ –koncentrácia plynu v čase t $[kg \cdot m^{-3}]$.

Krivka poklesu začína v bode, ktorý predstavuje počiatočnú koncentráciu C_0, a exponenciálne klesá, pokiaľ nie je dosiahnutá koncentrácia v okolnom vzduchu C_a [19]. Použitím rovnice (2.9) sa vypočíta krivka predpokladaného poklesu plynu. Parametre ako vonkajšia a počiatočná koncentrácia je možné ľahko zistiť, prípadne odmerať. Problémom výpočtu je stanovenie hodnoty intenzity vetrania. Intenzitu vetrania je však možné vypočítať pomocou exponenciálnej regresie tak, aby predpokladaná krivka čo najlepšie opisovala krivku nameraných hodnôt oxidu uhličitého. Ukážka prekrytia týchto kriviek je zobrazená na obrázku 2.1.



Obr. 2.1: Ukážka prekrytia krivky predpokladanej a nameranej koncentrácie oxidu uhličitého

Vzťah (2.9) je možné použiť len v prípade, že v miestnosti nie je žiaden zdroj emisii. Zistenú hodnotu intenzity vetrania je možné nazvať aj ako infiltrácia. Infiltrácia je typ výmeny vzduchu, kedy vzduch prúdi do budovy a z nej za pomoci netesností v obvodovom plášti budovy.

2.6 Hodnotenie vnútorného prostredia budov

Pri nedostatočnom vetraní dochádza k poškodzovaniu zdravia a k znižovaniu pocitovej pohody u ľudí, ktorí sa nachádzajú v nevetranom priestore. Podľa vyhlášky pre Českú republiku slúži oxid uhličitý ako ukazovateľ kvality vnútorného prostredia a jeho koncentrácia vo vnútornom ovzduší nesmie prekročiť hodnotu 1 500 ppm. Vyhláška ďalej definuje minimálnu intenzitu vymieňaného vonkajšieho vzduchu na $25\,m^3 \cdot h^{-1}$ pre jednu osobu a minimálnu intenzitu vetrania na $0,5\,h^{-1}$ pre obytnú miestnosť [28]. Pri štúdiu rôznych noriem je možné zistiť, že sa líšia v odporúčaných hodnotách a limitoch, ktoré sa týkajú okrajových podmienok pre vnútorné prostredie budov.

Požiadavky na množstvo čerstvého vzduchu sú dané aktivitou, emisiami vyprodukovanými budovou a zariadeniami v nej. Norma ČSN EN 16798-3 popisuje energetickú náročnosť budov a definuje štyri kategórie kvality vnútorného ovzdušia (IDA 1, IDA 2, IDA 3, IDA 4) na základe koncentrácie CO_2 nad koncentráciou vo vonkajšom prostredí [38]. Pre každú kategóriu sú definované doporučené hodnoty intenzity výmeny vzduchu, ktoré sa pohybujú v nefajčiarskej oblasti od $22\,m^3 \cdot h^{-1}$ a vo fajčiarskej od $43{,}2\,m^3 \cdot h^{-1}$ na jednu osobu [12]. V tabuľke 2.5 sú zobrazené jednotlivé kategórie a potrebné množstvo privádzaného čerstvého vzduchu na osobu. Čím chceme docieliť kvalitnejší vzduch v miestnosti, tým musí byť množstvo privádzaného vzduchu väčšie.

Kategória	**Kvalita vnútorného vzduchu**	**Koncentrácia CO_2 nad koncentráciou vo vonkajšom vzduchu** [ppm]		**Množstvo privádzaného vzduchu** [$m^3 \cdot h^{-1} \cdot osoba$]	
		Typický interval	**Implicitná hodnota**	**Nefajčiarska oblasť**	**Fajčiarska oblasť**
IDA 1	vysoká	<400	350	>54	>108
IDA 2	stredná	400–600	500	36–54	72–108
IDA 3	mierna	600–1 000	800	22–36	43,2–72
IDA 4	nízka	>1 000	1 200	<22	<43,2

Tabuľka 2.5: Základná klasifikácia kvality vnútorného ovzdušia podľa normy ČSN EN 16798-3

Popri kancelárskych, obytných a ďalších priestorov existujú aj školské zariadenia. Norma ČSN EN 16798-3 špecifikuje množstvo privádzaného vzduchu len na dospelú osobu a nedefinuje hodnoty pre žiakov v školských zariadeniach. Preto norma ČSN EN 15251 zohľadňuje aj produkované škodliviny samotnou budovou a definuje tak množstvo privádzaného vzduchu na $35\,m^3 \cdot h^{-1}$ pre deti v predškolských zariadeniach a $30\,m^3 \cdot h^{-1}$ pre žiakov v školských zariadeniach [12]. Pre Českú republiku je však záväzná vyhláška č. 20/2012 Sbírka zákonů, ktorá stanovuje množstvo privádzaného vzduchu do triedy na 20–30 $m^3 \cdot h^{-1}$ pre jedného žiaka.

Kapitola 3

Získavanie znalostí z databáz

Koncom 80. rokov a začiatkom 90. rokov minulého storočia došlo k výraznému zvyšovaniu kapacity úložných zariadení a zároveň sa začali objavovať nové databázové technológie. Okrem rozvoja existujúcich databázových technológií vznikali aj rôzne nové typy databáz, ako napríklad dátové sklady, multidimenzionálne databáze alebo databáze na webe. Rozvoj týchto technológii bol značne ovplyvnený výrazným nárastom dát, ktoré bolo potrebné ukladať.

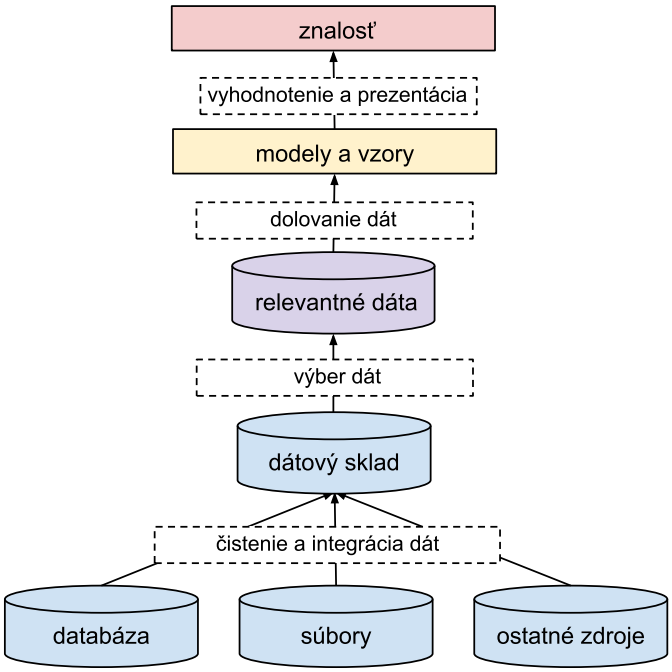
Množstvo dát, ktoré je potrebné ukladať, sa neustále zvyšuje. Oblasti ako výroba, veda, predaj a štátna správa sa potýkajú s problémom zahltenia veľkým množstvom dát. Motiváciou pre vznik novej oblasti v počítačových vedách, nazývanej získavanie znalostí z databáz (dolovanie dát, dolovanie z dát) boli rastúce objemy dát dostupné v elektronickej podobe, z ktorých bolo potrebné získať užitočné informácie a znalosti. Jedná sa o multidisciplinárny obor, čo znamená, že v sebe zahŕňa niekoľko samostatných oborov. Do získavania znalostí z databáz sa radia databázové technológie, strojové učenie, matematická štatistika ale aj rozpoznávanie vzorov, rôzne algoritmy, vizualizácie a ďalšie. Oblasť dolovania dát sa zameriava na extrakciu zaujímavých informácii, ktoré sú netriviálne (nedajú sa získať napríklad zložitým SQL dotazom) a nie sú na prvý pohľad viditeľné. Požadovaná informácia môže byť uložená v rozsiahlych databázach, na webe, v texte alebo v prúde dát.

Problematikou budovania dátových skladov sa začalo intenzívne zaoberať už v 90. rokoch minulého storočia. Dátový sklad je databáza, ktorá umožňuje analyzovať rozsiahle dáta. V týchto rokoch vznikli dva systémy, OLTP (On-Line Transaction Processing) a OLAP (On-Line Analytical Processing). Systém OLTP je orientovaný na zákazníka, zabezpečuje vykonávanie on-line transakcii a spracováva dotazy nad operačnou databázou. Rieši každodenné operácie, ktorými môžu byť predaj, naskladnenie alebo rezervácia tovaru. OLAP narozdiel od OLTP slúži k analýze dát a následnej reprezentácii výsledkov podľa požiadaviek. Typicky bývajú databáze pre oba systémy oddelené. Databáze pre OLTP obsahuje aktuálne dáta s potrebou riadenia súbežného prístupu. Databáza pre ukladanie historických dát, ktorá môže byť integrovaná z niekoľkých zdrojov, je určená pre OLAP systémy.

Získané informácie je možné uplatniť pri analýze nákupného košíka, trhu, detekcii podvodov (odľahlých hodnôt), predpovedi počasia alebo analýze textu. Špecifickým prípadom môže byť dolovanie znalostí z prúdu dát, ako je napríklad sieťová prevádzka. V poslednom období sa však s využitím získavania znalostí môžeme stretnúť aj v oblasti bioinformatiky a pri analýze biologických dát. Informácie uvedené v kapitole boli čerpané hlavne z [35].

3.1 Proces získavania znalostí

Získavanie znalostí z databáze je proces, ktorý sa skladá z určitých krokov, znázornených na obrázku 3.1. Niektoré kroky sa môžu opakovať, a tak môže byť dolovací proces obohatený o nové informácie z predchádzajúcich iterácii.



Obr. 3.1: Proces získavania znalostí

Jednotlivé kroky procesu získavania znalostí z databáze sú nasledujúce.

1. Čistenie dát – jedná sa o základnú operáciu na začiatku procesu. Cieľom čistenia dát je úprava dát, ktoré sú zašumené, nekonzistentné alebo v nich chýbajú niektoré položky.

2. Integrácia dát – dáta môžu pochádzať z rôznych zdrojov a môžu byť uložené v rôznych formátoch, čo môže viesť k nekonzistencii dát. Kroky čistenia a integrácie sa vykonávajú často spoločne. Dôvodom spoločného vykonávania je, že vyčistené dáta z rôznych zdrojov je potrebné niekde uložiť. Pre uloženie dát môže byť použitý dátový sklad alebo relačná databáza.

3. Výber dát – po úprave dát je potrebné vybrať dáta, ktoré sú pre riešenie úlohy relevantné. Pri získavaní znalostí z dát uložených v dátovom sklade vyberáme relevantné dimenzie. Pokiaľ pracujeme s relačnou databázou, vyberáme požadované tabuľky a relevantné stĺpce v nich.

4. Transformácia dát – úlohou transformácie dát je zjednotenie dát do takej formy, aby bolo získavanie znalostí z dát čo najmenej problematické. V tomto kroku sa dáta často sumarizujú, agregujú alebo normalizujú.

5. Dolovanie dát – predstavuje podstatu celého procesu získavania znalostí. Jedná sa o aplikáciu vybranej metódy a zvoleného algoritmu s určitými parametrami, ktorý vytvori vzory, respektíve modely.

6. Hodnotenie modelov a vzorov – dolovaním dát je možné získať veľké množstvo vzorov a modelov, pričom je potrebné určiť mieru zaujímavosti. Model alebo vzor je zaujímavý, ak je nový, zrozumiteľný a použiteľný.

7. Prezentácia znalostí – cieľom kroku je podať výsledky v zrozumiteľnej podobe pre bežného užívateľa alebo analytika. Výsledky môžu byť zobrazené pomocou grafov a ďalších vizualizačných techník.

3.2 Typy úloh pre dolovanie dát

Počas procesu získavania znalostí je dôležitým faktorom výber dolovacieho modelu, prípadne zvolenia skupiny modelov, ktoré vo výsledku dávajú lepší výsledok ako jednotlivé modely. Typy dolovacích úloh, ktoré je možné v kroku dolovania dát zvoliť, môžeme rozdeliť na deskriptívne a prediktívne modely. Deskriptívne modely popisujú dáta a ich štruktúru, hlavnými zástupcami sú:

- frekventované vzory,
- asociačné pravidlá,
- zhlukovanie.

Frekventované vzory sú vzory, ktoré sa v dátach vyskytujú často. Získaním frekventovaných vzorov môžeme nadobudnúť zaujímavé asociácie. Asociačné pravidlá vznikli v súvislosti s analýzou nákupného košíka, kde sa často analyzuje, aké produkty sa kupovali zároveň, prípadne aká skupina ľudí kupuje tieto položky. Posledná kategória deskriptívnych modelov je zhlukovanie, čo je proces, pri ktorom sa rozdeľujú dáta do tried (tzv. clusterov) na základe určitej podobnosti. Výsledné triedy obsahujú dáta, ktoré sú čo najviac podobné medzi sebou, a zároveň sa tieto dáta čo najviac odlišujú od dát z iných tried. Podobnosť objektov je možné určiť pomocou vzdialenostnej funkcie, ktorá môže byť založená na rozdelení, na hustote alebo na mriežke.

Opačným prípadom deskriptívnych modelov sú prediktívne modely, ktoré sa snažia popísať závislosti medzi jednotlivými atribútmi. Tieto modely sa usilujú skúmaním dát predpovedať chovanie v budúcnosti. Poprední zástupcovia týchto modelov sú z oblasti matematickej štatistiky a strojového učenia. Prediktívne modely môžeme rozdeliť na dve základné kategórie:

- klasifikácia,
- predikcia.

Klasifikácia umožňuje priradzovať dáta do tried na základe atribútov, ktorými sú analyzované dáta charakterizované. Výsledkom klasifikácie je model pre odhad triedy, do ktorej neznámy objekt patrí. Klasifikačný proces pozostáva z troch nižšie uvedených krokov:

1. trénovanie – rozborom trénovacej množiny je zostrojený klasifikačný model,
2. testovanie – po vytvorení klasifikačného modelu je potrebné vyhodnotiť jeho kvalitu, k čomu slúžia testovacie dáta,
3. aplikácia – nasadenie klasifikačného modelu na dáta s neznámou triedou.

Modely spadajúce do kategórie pre predikciu sa vyznačujú predpoveďou hodnoty určitej spojitej veličiny. Predpoveď hodnoty sa realizuje na základe atribútov analyzovaného záznamu. Príkladom predikcie môže byť výška platu po niekoľkých rokoch.

3.3 Klasifikačné a predikčné metódy

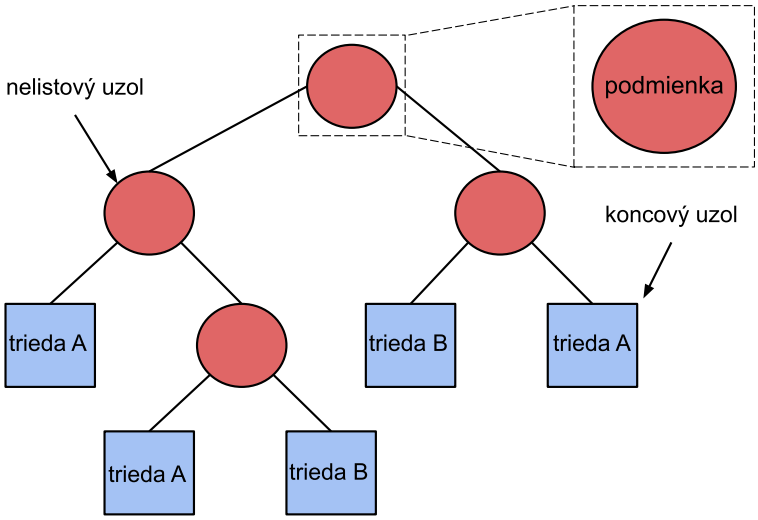
Cieľom klasifikačných a predikčných metód je nájdenie štruktúr v dátach, založených na vzájomných vzťahoch, a vytvorenie modelu alebo vzoru, ktorý pracuje samostatne bez trénovacích dát. V procese klasifikácie a predikcie dochádza k zhlukovaniu objektov do tried na základe hodnôt ich atribútov. Triedy následne obsahujú objekty s podobnými alebo rovnakými hodnotami atribútov. Objekty s odlišnými hodnotami atribútov sa naopak nachádzajú v rozdielnych triedach.

Ďalšou možnosťou pre klasifikáciu a predikciu môže byť kombinácia viacerých modelov a následné skupinové rozhodovanie. Príkladom môže byť metóda bagging, ktorá dáva váhu jednotlivým modelom, boosting, ktorý ohodnocuje klasifikátory podľa ich úspešnosti alebo stacking, ktorý sa sám učí, komu má ako veľmi dôverovať [11].

3.3.1 Rozhodovací strom

Rozhodovacie stromy (decision trees) patria medzi najpoužívanejšie metódy pre klasifikovanie dát. Dôvodom ich častého použitia je prehľadnosť a jednoduchá interpretovateľnosť výsledkov.

Rozhodovací strom je graf stromovej štruktúry. Skladá sa z nelistových a koncových uzlov, ako je zobrazené na obrázku 3.2. Nelistové uzly slúžia na testovanie hodnôt atribútov. Koncové uzly reprezentujú triedy, do ktorých môže byť objekt zaradený. Ďalej sa strom skladá z vetví, ktoré značia výstupy testov. Cesta od koreňa stromu ku koncovému uzlu tvori klasifikačné pravidlo. Takto je možné celý rozhodovací strom reprezentovať pomocou klasifikačných pravidiel.



Obr. 3.2: Ukážka jednoduchého rozhodovacieho stromu s klasifikáciou do dvoch tried

Vytvorenie rozhodovacieho stromu prebieha tak, že sa postupne vyberajú atribúty, ktoré maximálne odlišujú hodnoty. Z vybraného atribútu sa vytvori uzol a najčastejšie sa pome-

nuje názvom atribútu. Následne sa prejdu všetky hodnoty vybraného atribútu. Pre každú hodnotu vybraného atribútu sa vytvori vetva z posledne vytvoreného uzla. Ďalej je potrebné vytvoriť podmnožinu vzoriek z trénovacej množiny, pre ktoré platí, že sa ich hodnota atribútu rovná hodnote atribútu v poslednom vytvorenom uzle. Ak je podmnožina prázdna, pripojí sa koncový uzol s najbežnejšou triedou, inak sa rekurzívne vytvori podstrom, ktorý sa pripojí k danej vetve.

Najdôležitejšou časťou vytvorenia stromu je výber atribútu, ktorý maximálne odlišuje hodnoty v danom uzle. Pre výber vhodného atribútu je potrebná funkcia, ktorá splňuje nasledujúce kritéria. Čím je hodnota atribútu menej pravdepodobná, tým nesie väčšie množstvo informácie. Ďalším kritériom je, aby hodnota atribútu s výskytom s pravdepodobnosťí jedna niesla nulovú informáciu. Zjednodušene povedané chceme funkciu, ktorá pri rovnakej hodnote atribútu objektu v trénovacích dátach bude nadobúdať nulovú hodnotu. Tieto kritéria splňuje funkcia $Info(A) = -log_2(p(A))$, kde $p(A)$ je pravdepodobnosť výskytu hodnoty atribútu A. Priemerné množstvo informácie vybraného atribútu môžeme vypočítať podľa nižšie uvedenej rovnice (3.1), pričom hľadáme atribút A, ktorý dosahuje čo najnižšiu hodnotu entropie

$$Info(A) = -\sum_{i=1}^{m} p_i \cdot log_2(p_i). \tag{3.1}$$

Hodnota p_i je pravdepodobnosť, že náhodne zvolený objekt z trénovacej množiny je zaradený do triedy i a môže byt vypočítaná ako

$$p_i = \frac{s_i}{|S|}. \tag{3.2}$$

Hodnota s_i značí počet prvkov trénovacej množiny, ktoré sú zaradené do triedy i, $|S|$ označuje celkový počet prvkov v trénovacej množine.

Rozhodovací strom vyžaduje len diskrétne hodnoty. V prípade, že hodnoty atribútov nadobúdajú spojitý charakter, je potrebné nájsť čo najlepšiu deliacu hodnotu. Rozdelenie atribútu, ktorý nadobúda hodnôt spojitého charakteru, vykonáme tak, že najprv zoradíme hodnoty a potom aplikujeme výraz (3.3) na výpočet deliacej hodnoty

$$splitpoint_i = \frac{a_i + a_{i+1}}{2}. \tag{3.3}$$

Následne vypočítame priemerné množstvo informácie atribútov pomocou rovnice (3.1) a vyberieme také rozdelenie, aby hodnota priemerného množstva informácie bola čo najmenšia.

Existuje niekoľko algoritmov pre generovanie rozhodovacích stromov. Najznámejšie algoritmy sú ID3 a C4.5. ID3 využíva pre hľadanie priemerného množstva informácie vzťah (3.1). Rovnica však nezohľadňuje počet hodnôt atribútu. V prípade, že by atribút obsahoval unikátne hodnoty, bol by uprednostnený oproti ostatným atribútom. Riešením je metóda C4.5, ktorá zohľadňuje počet hodnôt atribútu. Existujú ešte ďalšie metriky pre výber atribútu, ktorý čo najlepšie rozdeľuje hodnoty do tried a niektoré z nich budú podrobnejšie popísané v kapitole 4.

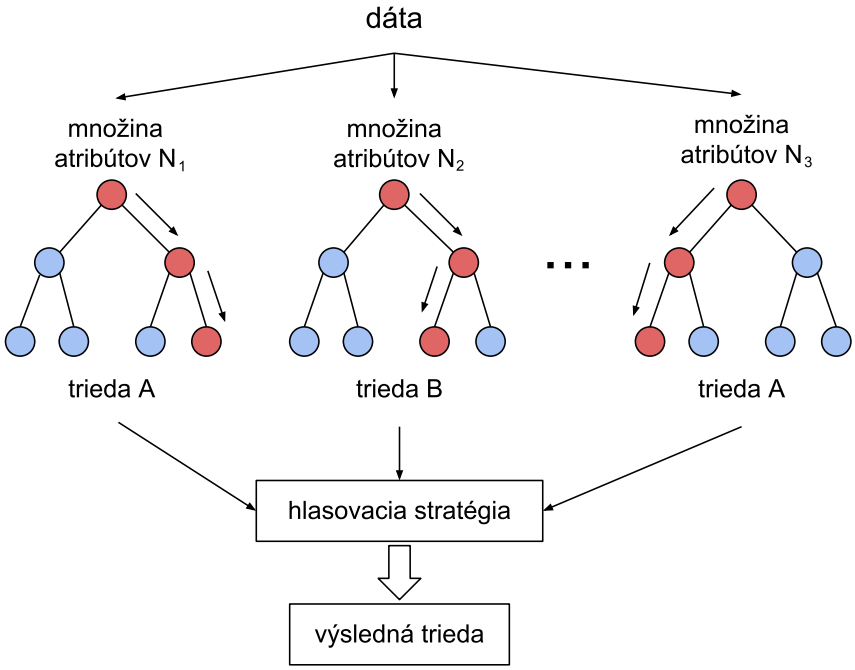
Výsledné rozhodovacie stromy môžu byť zložité a môžu dosahovať nižšiu presnosť. Dôvodom zložitosti môže byť napríklad šum v dátach, ktorý zapríčiní vznik nadbytočných vetví. Riešením je orezanie stromu od vetví, ktoré môžu zhoršovať celkovú presnosť. Orezanie vetví stromu môžeme rozdeliť na prepruning a postpruning. Metóda prepruning orezáva

vetvy stromu počas jeho vytvárania. Postpruning metóda oreže vetvy stromu až po jeho vytvorení. Metóda postpruning je výpočetne náročnejšia ako metóda prepruning, ale dosahuje lepšie výsledky.

3.3.2 Náhodný les

Náhodný les, tiež označovaný ako random forest, je metóda, ktorá pre klasifikáciu a predikciu využíva skupinu rozhodovacích stromov. Každý podstrom hrá rolu slabého klasifikátora, pričom ich výsledky sú vstupom do hlasovacej stratégie, ktorá rozhodne o celkovom výsledku, ako je zobrazené na obrázku 3.3. Hlasovacia funkcia môže vybrať výslednú triedu na základe najvyššej akumulovanej dôvery alebo najčastejšie predpokladanej triedy. Takouto kombináciou slabých klasifikátorov je možné vytvoriť silný klasifikátor, ktorý bude dobre fungovať aj na rozsiahlych dátach [23].

Jednotlivé rozhodovacie stromy sa konštruujú obdobným spôsobom ako bolo uvedené v podsekcii 3.3.1. Jediným rozdielom je, že pre každý slabý klasifikátor je náhodne vybraná podmnožina zo všetkých atribútov, čím dochádza k redukcii dimenzionality.



Obr. 3.3: Ukážka náhodného lesa

V prípade náhodného lesa nedochádza k preučeniu, ako to často býva v prípade klasického rozhodovacieho stromu. Preučenie je spôsobené veľmi špecifickým prispôsobením pravidiel na konkrétne dáta. Použitie preučeného rozhodovacieho stromu pre klasifikáciu iných dátových sad je potom nevhodné. Nespornou výhodou náhodného lesa je možnosť distribuovaného výpočtu, kedy každý strom pracuje s redukovaným počtom dát, čím sú znížené pamäťové nároky pri veľkých dátových sadách.

3.3.3 Bayesovská klasifikácia

Bayesovská klasifikácia je na rozdiel od rozhodovacích stromov založená na matematickej štatistike. Počíta s akou pravdepodobnosťou bude vybraný objekt patriť do každej triedy. Objekt je zaradený do triedy, pre ktorú je najväčšia pravdepodobnosť, že sa tu tento objekt vyskytuje. Podstatou Bayesovskej klasifikácie je Bayesov vzorec, ktorý je odvodený z výpočtu podmienenej pravdepodobnosti. Bayesov vzorec má tvar

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}. \tag{3.4}$$

Nech $X = (x_1, x_2, ..., x_n)$ reprezentuje hodnoty atribútov objektu X, ktorý má byť zaradený do jednej z tried $C_1, C_2, ..., C_n$. $P(C_i|X)$ udáva, aká je pravdepodobnosť, že prvok X patrí do triedy C_i, za predpokladu, že známe hodnoty atribútov. Takto vieme zistiť pravdepodobnosti príslušnosti objektu do všetkých tried. Cieľom výpočtu je zistiť hodnotu i, pre ktorú je hodnota $P(C_i|X)$ najvyššia. Hodnota $P(C_i|X)$ je maximálna práve vtedy, keď bude maximálny výraz $P(X|C_i) \cdot P(C_i)$. Dôvodom je, že hodnota $P(X)$ je pre vybraný objekt X konštantná.

Pravdepodobnosť, že vybraný prvok patrí do triedy C_i, získame použitím vzorca

$$P(C_i) = \frac{|s_i|}{|S|}. \tag{3.5}$$

Hodnota $|s_i|$ značí počet objektov, ktoré sa klasifikovali do triedy C_i, a $|S|$ označuje počet všetkých prvkov. Pravdepodobnosť, že vybraný prvok z triedy C_i bude mať zhodné atribúty s prvkom X je daná vzťahom

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i). \tag{3.6}$$

Pravdepodobnosť je definovaná ako súčin pravdepodobností, že prvok X s danými hodnotami atribútov x_k, kde $k = 1, \ldots n$ bude zaradený do triedy C_i, pričom n je počet atribútov.

Bayesovská klasifikácia pri výpočte $P(X|C_i)$ rozlišuje, či sa jedná o spojitú alebo o diskrétnu hodnotu atribútu. Výpočet $P(X|C_i)$ pre diskrétne hodnoty je daný vzťahom

$$P(X|C_i) = \frac{|s_{ik}|}{|s_i|}. \tag{3.7}$$

Hodnota s_{ik} určuje počet objektov z trénovacej množiny zaradených do triedy C_i, pričom hodnota jeho k-tého atribútu odpovedá hodnote x_k.

Výpočet $P(X|C_i)$ pre hodnoty spojitého charakteru je daný vzťahom

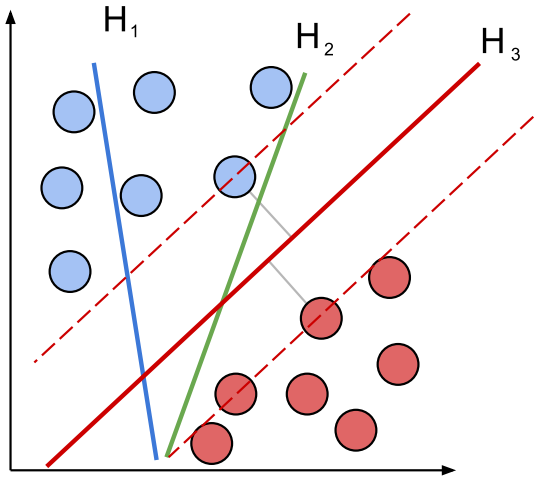
$$P(x_k|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}) = \frac{1}{\sqrt{2\pi\sigma_{c_i}}} e^{(-\frac{(x_k-\mu_{C_i})^2}{2\sigma_{c_i}^2})}. \tag{3.8}$$

Funkcia $g(x_k, \mu_{C_i}, \sigma_{C_i})$ označuje Gaussovo normálne rozloženie atribútu x_k, podľa ktorého sa bude klasifikovať do triedy C_i, μ_{C_i} značí priemernú hodnotu triedy C_i a σ_{C_i} označuje smerodatnú odchýlku triedy C_i.

3.3.4 Algoritmus podporných vektorov

Algoritmus podporných vektorov (Support vector machines – SVM) je používaný pre predikčné a klasifikačné úlohy. SVM metóda sa snaží nájsť nadrovinu, ktorá v priestore atribútov rozdelí trénovacie dáta do dvoch tried, pričom sa môže jednať o lineárne alebo nelineárne rozdelenie. Ak sa jedná o nelineárne rozdelenie, tak sa metóda často označuje ako SVM kernel metóda. Dôležitou súčasťou algoritmu podporných vektorov je jadrová transformácia priestoru atribútov do priestoru atribútov, ktorý má typicky vyššiu dimenzionalitu. Jadrová transformácia dokáže previesť pôvodne lineárne neseparovateľné atribúty do priestoru, kde je možné nájsť nadrovinu, ktorá umožní atribúty lineárne separovať [5].

V priebehu vytvárania SVM modelu je nájdená nadrovina. Cieľom je nájsť nadrovinu tak, aby medzi separovanými (diskriminovanými) atribútmi vznikol, čo najväčší pás (maximal margin alebo functional margin) [5]. Čím je možné nájsť väčší pás, tým bude menšia chybovosť pri klasifikácii a predikcii. Tento poznatok je zobrazený na obrázku 3.4, kde nadrovina H_1 neoddeľuje dve triedy, nadrovina H_2 oddeľuje dve triedy ale s malým pásom a nadrovina H_3 oddeľuje dve triedy s maximálnym možným pásom.



Obr. 3.4: Možnosti oddelenia atribútov dvoch tried pomocou SVM metódy

Na obrázku 3.4 sú tiež zobrazené podporné vektory, ktoré pretína čiarkovaná čiara. Podporné vektory, ktoré sa nachádzajú, čo najbližšie hranici, slúžia na popis nadroviny.

SVM model je možné použiť ako lineárny klasifikátor zaraďujúci dáta do dvoch tried, ktoré sa dajú lineárne separovať. Vstupom klasifikátoru sú trénovacie dáta, ktoré obsahujú dvojice $(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)$, kde $\vec{x}_1, \ldots, \vec{x}_n$ sú p-dimenzionálne vektory reálnych čísiel a $y_1, \ldots, y_n$ značí ich zaradenie do jednej z tried, a to v podobe hodnôt -1 a +1. Potom nadrovinu, ktorá oddeľuje dve triedy, je možné popísať vzťahom $0 = \vec{w} \cdot \vec{x} - b$, kde $\vec{w}$ je normála nadroviny a b označuje bias. Normálu nadroviny je možné využiť pre výpočet šírky pásma $\frac{2}{||\vec{w}||}$ [5]. Rovnice pre klasifikáciu sú

$$\vec{w} \cdot \vec{x}_n + b \leq -1 \quad pre\, y_n = -1, \tag{3.9}$$

$$\vec{w} \cdot \vec{x}_n + b \geq +1 \quad pre\, y_n = +1. \tag{3.10}$$

3.3.5 Umelé neurónové siete

Umelé neurónové siete patria medzi často využívané metódy klasifikácie a predikcie. Tieto siete sú založené na prepojení jednoduchých rozhodovacích blokov (umelých neurónov) do jednej veľkej siete. Inšpiráciou pre umelé neurónové siete je model ľudského mozgu, ktorý obsahuje neuróny pospájané axónmi. Informácie uvedené v podsekcii boli čerpané z [29].

Základom každej neurónovej siete je perceptron. Perceptron umožňuje lineárnu klasifikáciu do dvoch tried na základe kombinácie množiny binárnych vstupov $\{x_1, x_2, ..., x_n\}$ s vektorom váh $\{w_1, w_2, ..., w_n\}$. Vstupy perceptrónu sa najprv vynásobia príslušnými váhami a následne sa sčítajú, pričom vznikne vážený súčet vstupov, kde N je počet hodnôt

$$\sum_{n=0}^{N} w_i \cdot x_i. \tag{3.11}$$

Ak je výsledok váženého súčtu vstupov menší než prahová hodnota (threshold), výsledok je nula, v opačnom prípade je výsledok jedna.

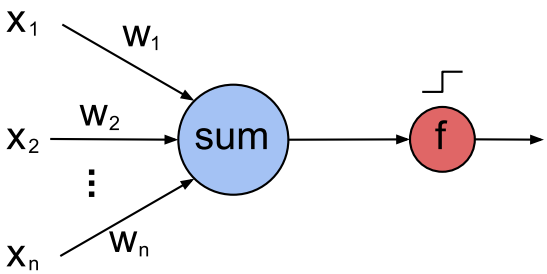
V praxi sa však používa upravený perceptrón, kde sa do neho zavádza systémová chyba, ktorá sa anglicky označuje bias. Bias sa označuje písmenom b, pričom platí, že $b = -threshold$. Výstup perceptrónu sa potom spočíta ako skalárny súčin vektorov váh a vstupných hodnôt, kedy sa k tomuto súčinu pričíta ešte hodnota biasu

$$y = w \cdot x + b. \tag{3.12}$$

Celkový výstup je nula, ak je vypočítaná hodnota menšia ako nula, v opačnom prípade je výsledok jedna.

Spomenuté varianty perceptrónu majú nevýhodu v tom, že je potrebné zadávať váhy a biasy ručne. Aby mohla umelá neurónová sieť konkurovať ostatným klasifikačným a predikčným metódam, je potrebné nastavovať váhy a biasy algoritmicky. Cieľom je nájsť algoritmus, ktorý bude schopný na základe vstupných vektorov a výstupov odvodiť ideálne váhy a biasy. Pomocou vypočítaných váh a biasov sa vie umelá neurónová sieť veľmi dobre naučiť klasifikovať trénovacie dáta. Ručné zadávanie váh a biasov je možné eliminovať pomocou zavedenia aktivačnej funkcie, ktorá sa označuje písmenom σ. Vstupom aktivačnej funkcie je súčet biasu a skalárneho súčinu vektoru váh a vstupných hodnôt

$$y = \sigma(w \cdot x + b). \tag{3.13}$$



Obr. 3.5: Schéma perceptrónu neurónovej siete

Schéma jedného neurónu, ktorý aktivačnú funkciu využíva, je uvedené na obrázku 3.5. Výstup aktivačnej funkcie sa nachádza v určitom intervale, ktorý umožní zmierniť vplyv váh na celkový výsledok umelej neurónovej siete. Najčastejšie sa ako aktivačná funkcia používa sigmoid, jej výstup spadá do intervalu $(0; 1)$, alebo hyperbolický tangens s výstupmi v intervale $(-1; 1)$ a ďalšie.

Neurónová sieť sa skladá z troch vrstiev. Na začiatku umelej neurónovej siete sa nachádza vstupná vrstva, do ktorej vstupuje vektor atribútov. Výstupná vrstva sa nachádza na konci neurónovej siete a slúži na klasifikovanie do určených tried. Medzi týmito vrstvami sa môže nachádzať niekoľko vrstiev, označovaných ako skryté vrstvy, ktoré sa využívajú pri hlbokom učení.

3.4 Nástroje na dolovanie dát

Pre implementáciu vyššie spomenutých algoritmov strojového učenia ale aj pre implementáciu redukcie dimenzionality existuje celá rada nástrojov. Nástroje sa líšia v počte podporovaných metód a možnostiach ich nastavenia. Niektoré nástroje ponúkajú jednoduchú prácu s dátovými sadami (ich načítanie, rozdelenie, spájanie) a podporujú mnoho spôsobov vizualizácie výsledkov. Medzi najčastejšie používané nástroje na dolovanie dát patrí RapidMiner, Weka, KNIME, R, Orange a ďalšie [27].

RapidMiner je softwérová platforma, ktorá bola vyvinutá spoločnosťou Rapid Miner. Softwér poskytuje prostredie pre strojové učenie, dolovanie dát a textu, alebo prediktívne a obchodné analýzy. Pre návrh zložitých problémov je možné využiť veľké množstvo ľubovoľne vnorených operátorov. Každý operátor vyžaduje určité vstupy, ktoré daným spôsobom spracuje a výsledky spracovania sprístupní na výstupe. Výstupy môžu byť ďalej predané na vstup ďalšieho operátora. RapidMiner môže byť použitý pre výskum, vzdelávanie, trénovanie, rýchly vývoj prototypovej aplikácie alebo obchodné a priemyselné aplikácie. Jednou z výhod nástroja je dostupnosť veľkého počtu metód pre integráciu, transformáciu, analýzu, modelovanie a vizualizáciu dát. Ďalej ponúka mnoho možností, ako vyberať atribúty, alebo detekovať odľahlé hodnoty [27].

Weka (Waikato Environment for Knowledge Analysis) je open source sada programov pre strojové učenie. Umožňuje modelovanie dát, ich analýzu a vizualizáciu. Pre Weku bol vytvorený vlastný textový formát ARFF (Attribute Relationship File Format), ktorý je podobný CSV súboru. Obsahuje atribúty, dodatočné informácie pre klasifikáciu a samotné dáta. Na začiatku súbora sú popísané jednotlivé stĺpce, ich typy a rozsah ich hodnôt. Za popisom dát nasledujú samotné dáta oddelené čiarkou. Weka podporuje rôzne druhy metód pre klasifikáciu, predikciu, zhlukovanie a redukciu dimenzionality [3].

Orange je softwér vhodný pre strojové učenie a dolovanie dát. Poskytuje komponenty pre predspracovanie dát, ohodnotenie a filtrovanie atribútov, modelovanie a jeho vyhodnotenie. Nástroj bol vyvinutý na fakulte počítačových a informačných vied univerzity v Slovinsku, za spolupráce s open source komunitou. Umožňuje robiť rýchle rozhodnutia za použitia rýchleho porovnania a analýzy dát [32].

Kapitola 4

Redukcia dimenzionality

V rámci kroku dolovania dát je potrebné určiť nie len typ dolovacej úlohy, ale aj relevantnú množinu atribútov, tiež nazývanú množina rysov (feature vector). Pri veľkom množstve atribútov môže dochádzať k tzv. prekliatiu dimenzionality (the curse of dimensionality). Jedná sa o koncept poukazujúci na jav, kedy sú dáta viacej rozptýlené vo vysoko dimenzionálnom priestore, čo nepriaznivo ovplyvňuje algoritmy pracujúce s malým počtom dimenzii [20]. Pre štatisticky spoľahlivé výsledky dát vo vysoko dimenzionálnom priestore je potrebné veľké množstvo dát, čo má za následok zvyšujúce sa požiadavky na pamäťový priestor a cenu výpočtu celej analýzy [35].

Najjednoduchší spôsob výberu najlepšej podmnožiny atribútov je vytvoriť všetky podmnožiny, ktoré by sa mohli použiť na trénovanie, a následne vybrať takú podmnožinu, ktorá dosahuje najvyššiu úspešnosť. Takéto hľadanie podmnožiny je výpočetne náročné aj pre malý počet atribútov. Počet všetkých podmnožin, ktoré by bolo potrebné otestovať [24], je kombinačné číslo dané vzťahom

$$C(d, m) = \frac{d!}{m!(d-m)!}, \tag{4.1}$$

kde jednotlivé parametre $C(d, m)$, d a m znamenajú:

$C(d, m)$ – počet podmnožin na otestovanie,

d – počet atribútov,

m – mohutnosť podmnožiny.

Počet podmnožin na otestovanie je možné podľa vyššie uvedeného vzťahu vypočítať len za predpokladu, že poznáme mohutnosť podmnožiny m, čo bežne nevieme. Častejšie nás zaujímajú podmnožiny, pričom ich mohutnosti sú nanajvýš rovné mohutnosti podmnožiny m [24]. Podmnožin je oveľa väčšie množstvo, ako v prípade, keď poznáme mohutnosť podmnožiny

$$\sum_{i=1}^{m} C(d, m). \tag{4.2}$$

Redukcia dimenzionality je najsilnejším nástrojom pre riešenie problémov s prekliatím dimenzionality a otestovaním veľkého množstva podmnožin.

Metódy redukcie dimenzionality je možné rozdeliť na dve hlavné skupiny [35]:

- extrakcia atribútov (feature extraction),
- výber atribútov (feature selection).

4.1 Extrakcia atribútov

Extrakcia atribútov je technika, počas ktorej sa pôvodné atribúty transformujú do nového priestoru atribútov. Rozmer novej dimenzie je rovnaký alebo menší ako pôvodný rozmer.

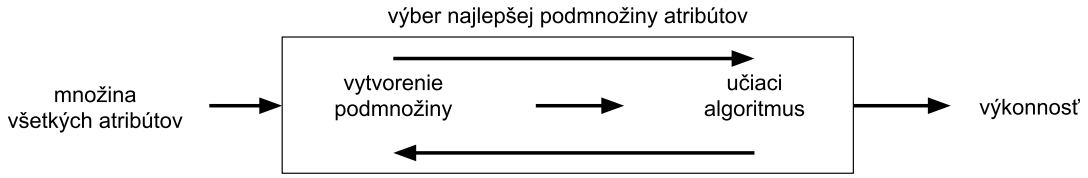
Hlavným predstaviteľom metód pre extrakciu atribútov je analýza hlavných komponent (Principal Component Analysis). Jej cieľom je zjednodušenie popisu skupiny lineárne závislých atribútov. Analýzu hlavných komponent je možné tiež popísať ako lineárnu transformáciu pôvodných atribútov na nové nekorelované atribúty, ktoré nazývame hlavné komponenty. Každú hlavnú komponentu charakterizuje jej miera variability. Tieto komponenty sú následne zoradené podľa ich hodnoty variability od najväčšej k najmenšej [24].

4.2 Výber atribútov

Metóda výberu atribútov redukuje dimenzionalitu pomocou hľadania minimálnej podmnožiny atribútov na základe ďalšieho spracovania, alebo kritéria výberu. Cieľom výberu je nájdenie takej podmnožiny atribútov, ktorá sa použijú pre dolovanie a zároveň sa vylúčia málo relevantné alebo redundantné atribúty. Metódy výberu atribútov je možné rozdeliť na wrapper, filtračné a embedded metódy [16, 20].

4.2.1 Wrapper metódy

Názov tejto metódy vychádza z toho, že metóda obaľuje použitý algoritmus strojového učenia, ktorý následne slúži k vyhodnoteniu vybranej podmnožiny atribútov. Najskôr je uvažovaná množina všetkých atribútov, z ktorej sa postupne vytvárajú podmnožiny. Kritérium (napríklad úspešnosť, klasifikačná chyba) modelu, ktorý pracuje s danou podmnožinou, sa použije ako vzdialenostná funkcia pre výber najlepšej podmnožiny. Nakoniec je model otestovaný na testovacích dátach, ktoré neboli použité v procese trénovania, a určí sa výsledná presnosť klasifikátora [16]. Vizualizácia postupnosti krokov wrapper metódy je zobrazená na obrázku 4.1.



Obr. 4.1: Vizualizácia krokov wrapper metódy

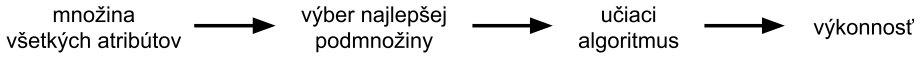
Takýto prístup hrubou silou je výpočetne náročný, ale poskytuje najlepšie výsledky pre klasifikačný model použitý vo wrapper metóde. Použitím správnej stratégie prehľadávania priestoru atribútov je možné eliminovať veľkú časť irelevantných podmnožin.

Príkladom môžu byť metódy dopredného výberu a spätnej eliminácie [16]:

- dopredný výber (Forward Selection) – iteračná metóda, ktorá postupne v každej iterácii pridá atribút s cieľom zlepšiť úspešnosť zvoleného modelu,
- spätná eliminácia (Backward Elimination) – iteračná metóda, ktorá na začiatku uvažuje všetky atribúty a potom v každej iterácii odstráni najmenej relevantný atribút.

4.2.2 Filtračné metódy

Filtračné metódy sa používajú počas predspracovania dát. Princíp metód, ktoré patria do tejto skupiny, je založený na štatistickom ohodnotení každého atribútu. Na základe ohodnotenia všetkých atribútov sa vyberie podmnožina atribútov s najlepším ohodnotením. Metódy nie sú závislé od použitého algoritmu pre strojové učenie, ako je to v prípade wrapper a embedded metód. Filtračné metódy bývajú z hľadiska časovej náročnosti menej náročné než ostatné metódy [16]. Jednotlivé kroky filtračnej metódy sú zobrazené na obrázku 4.2.

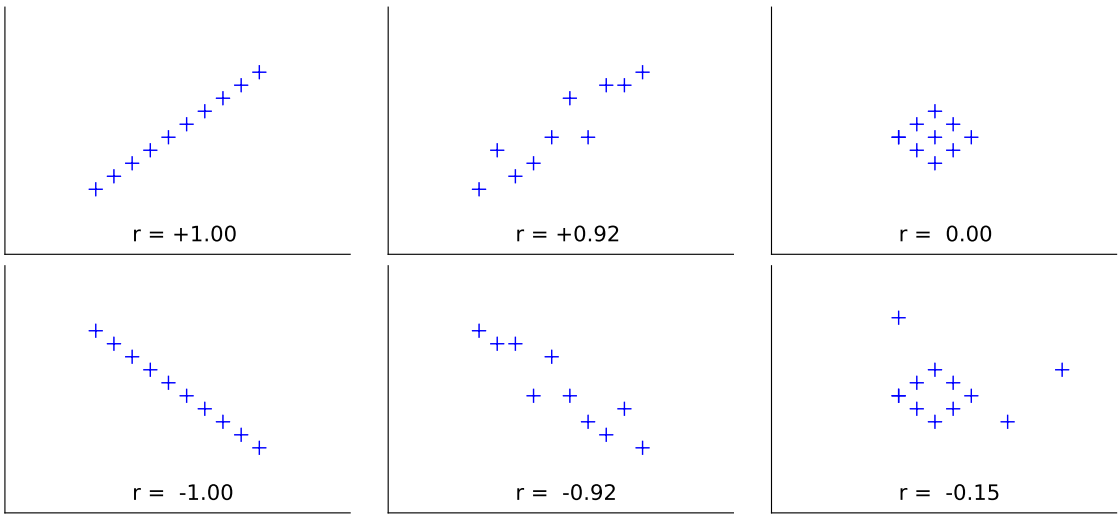


Obr. 4.2: Vizualizácia krokov filtračnej metódy

Filtračné metódy môžeme rozdeliť na metriku založenú na korelácii (Pearsonov korelačný koeficient), štatistickú metriku (Chí-kvadrát test), informačne-teoretickú metriku (entropia, informačný zisk) a ostatné metriky (Gini index) [20].

Pearsonov korelačný koeficient

Pearsonov korelačný koeficient slúži na určenie veľkosti lineárnej závislosti medzi hodnotami dvoch atribútov. Korelačný koeficient nadobúda hodnoty z intervalu $\langle 1; -1 \rangle$, ako je zobrazené na obrázku 4.3.



Obr. 4.3: Grafické zobrazenie závislosti hodnôt atribútov a ich korelačné koeficienty

Pokiaľ je hodnota korelačného koeficientu rovná 1, tak medzi atribútmi platí vzťah priamej úmery (pozitívne korelované atribúty). Ak je hodnota rovná -1, platí vzťah nepriamej

úmery (negatívne korelované atribúty). V prípade, že korelačný koeficient nadobúda hodnoty 0, sú atribúty nezávislé [35]. Personov korelačný koeficient je možné vypočítať pomocou vzťahu

$$r = \frac{\sum_{i=1}^{n} (x_i y_i) - n\overline{X}\,\overline{Y}}{(n-1)\sigma_x \sigma_y}, \tag{4.3}$$

kde jednotlivé parametre r, n, $\overline{X}$, $\overline{Y}$, σ_x a σ_y znamenajú:

r – Pearsonov korelačný koeficient,

n – počet n-tic,

$\overline{X}, \overline{Y}$ – priemer atribútov x a y,

σ_x, σ_y – smerodatná odchýlka atribútov x a y.

Treba však brať do úvahy aj to, že korelácia neimplikuje kauzalitu [35]. Napríklad zistená korelácia atribútu počet škôl a počet nemocníc v meste nejde interpretovať tak, že je v meste viacej nemocníc, pretože je v meste viacej škôl. Hlavným atribútom, ktorý ovplyvňuje koreláciu atribútov, môže byť v tomto prípade počet obyvateľov udávajúci veľkosť mesta.

Chí-kvadrát test

Chí-kvadrát test, často označovaný len ako χ^2 alebo test dobrej zhody, je používaný k zisteniu pravdepodobnosti korelácie, alebo asociácie medzi kategorickými atribútmi na základe frekvenčného rozloženia [35]

$$\chi^2 = \sum_{i=1}^{c} \sum_{j=1}^{r} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}, \tag{4.4}$$

kde jednotlivé parametre χ^2, o_{ij} a e_{ij} znamenajú:

χ^2 – test dobrej zhody,

o_{ij} – pozorovaný počet výskytov,

e_{ij} – očakávaný počet výskytov.

Výsledkom testu dobrej zhody je, ako veľmi sa líši očakávaná frekvencia výskytov od nameranej frekvencie. V prípade, že test dobrej zhody dosahuje malé hodnoty, sú atribúty nezávislé. Vysoké hodnoty naopak ukazujú závislosť atribútov [35].

Veľkou výhodou tejto metódy je jej jednoduchosť, pričom nekladie žiadne požiadavky na to, ako by mali byť dáta rozdelené. Nevýhodou tejto metriky je, že hodnoty atribútov nesmú na sebe byť závislé a nemôžu nadobúdať záporné hodnoty.

Entropia

Entropia patrí medzi informačne-teoretické metriky, ktoré ohodnocujú atribúty na základe ich dôležitosti. Princíp informačne-teoretickej metriky je postavený na entropii, ako miere neurčitosti systému v spracovávaných atribútoch [20]

$$H(Y) = -\sum_{y \in Y} p(y) log_2(p(y)), \tag{4.5}$$

kde jednotlivé parametre $H(Y)$ a $p(y)$ znamenajú:

$H(Y)$ – entropia pre atribút Y,

$p(y)$ – pravdepodobnosť výskytu hodnoty atribútu y.

V prípade, že sú hodnoty rovnomerne rozložené do tried, dosiahneme najvyššiu entropiu. Ak sa hodnoty spracovaného atribútu nachádzajú len v jednej triede, tak je entropia nulová.

Informačný zisk

Metódu informačného zisku môžeme zaradiť do kategórie informačne-teoretických metrík, ktoré sú založené na entropii. Informačný zisk je skóre, ktoré určuje väzbu medzi atribútom a triedou [20]. Pre výpočet informačného zisku je potrebné vedieť vypočítať entropiu premennej Y za predpokladu pozorovania hodnôt premennej X

$$H(Y) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) log_2(p(y|x)). \tag{4.6}$$

Výsledný vzťah pre výpočet informačného zisku je následne možné definovať ako zníženie entropie daného atribútu za predpokladu pozorovania iného atribútu [20]

$$InformationGain = H(Y) - H(Y|X) = H(X) - H(X|Y). \tag{4.7}$$

Informačný zisk atribútu Y, za predpokladu pozorovania atribútu X, je rovnaký, ako informačný zisk atribútu X, za predpokladu pozorovania atribútu Y. Čím je výsledná hodnota vyššia, tým je atribút užitočnejší. Nevýhodou tejto metriky je, že je ovplyvnená počtom hodnôt atribútov.

Gini index

Pre výber atribútov je možné použiť mnoho metrík, ktoré nepatria do žiadnej spomenutej kategórie. Medzi tieto metriky zaraďujeme aj Gini index, ktorý pre každý atribút vypočíta hodnotu Gini indexu, potom hodnoty zoradí a vyberie atribúty s nejmenšou hodnotou [20]. Pre výpočet metriky sa používa vzťah

$$G(S) = 1 - \sum_{i=1}^{m} p(C_i|X)^2, \tag{4.8}$$

kde jednotlivé parametre $G(S)$ a $p(C_i|X)$ znamenajú:

$G(S)$ – Gini index,

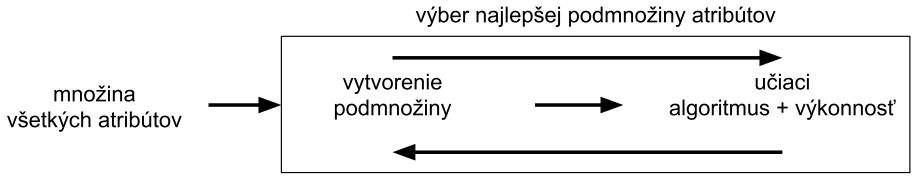
$p(C_i|X)$ – pravdepodobnosť, že atribút X patrí do triedy C_i za predpokladu, že poznáme hodnoty atribútu X.

Hodnota Gini indexu je minimálna v prípade, že hodnoty atribútov patria do jednej triedy, a maximálna v prípade, že hodnoty atribútov sú rozdelené medzi triedami.

4.2.3 Embedded metódy

Embedded metódy kombinujú výhody filter a wrapper metód. Ako embedded metódy sú označované metódy, ktoré sú priamo súčasťou algoritmu pre strojové učenie. Filtračné metódy v procese výberu nezahŕňajú fázu učenia, pretože sa používajú v procese predspracovania. Wrapper metódy zisťujú kvalitu danej podmnožiny atribútov pomocou prediktora, pričom je vybraná podmnožina závislá na použitom prediktore [16].

Embedded metódy pristupujú k výberu najlepšej podmnožiny iným spôsobom, ako je zobrazené na obrázku 4.4.



Obr. 4.4: Vizualizácia krokov embedded metódy

Výber najlepšej podmnožiny a proces učenia sa vykonáva v jednom kroku, kedy sa vyberú relevantné atribúty prislúchajúce pre zvolený klasifikačný algoritmus. Príkladom embedded metód môže byť rozhodovací strom, ktorý vyberá vhodné atribútu na základe metriky, akou je napriklad informačný zisk alebo Gini index.

Kapitola 5

Merania oxidu uhličitého

Meranie koncentrácie oxidu uhličitého vo vnútorných priestoroch je realizované hlavne z dôvodu zistenia kvality vzduchu v tomto priestore, pretože CO_2 slúži ako ukazovateľ kvality ovzdušia. Vyhovujúcu kvalitu vnútorného ovzdušia je možné udržať správnym vetraním a dodržiavaním limitov pre koncentráciu CO_2 v budovách a bytových priestoroch. Dodržiavanie limitov znižuje riziko nepohody osôb, vzniku chorôb súvisiacich s nadmernou koncentráciou CO_2 a neznižuje efektivitu ľudí pri práci. Na meranie koncentrácie CO_2 môžeme použiť dve kategórie senzorov, a to snímače oxidu uhličitého a snímače zmesi plynov [34].

Prínosom merania koncentrácie CO_2 môže byť riadenie núteného vetrania, ktoré podľa aktuálnej úrovne CO_2 upraví množstvo privádzaného čerstvého vzduchu do priestorov (spätnoväzobné nútené vetranie).

Okrem spätnoväzobného núteného vetrania je možné namerané hodnoty ukladať a následne využiť pre analýzu. Analýzou dát môžu byť zistené rôzne vzory, závislosti a trendy, ktoré sa môžu opakovať s určitou frekvenciou. Pre zber dát zo senzorov, ich ukladanie a následné získanie bol zvolený BeeeOn systém.

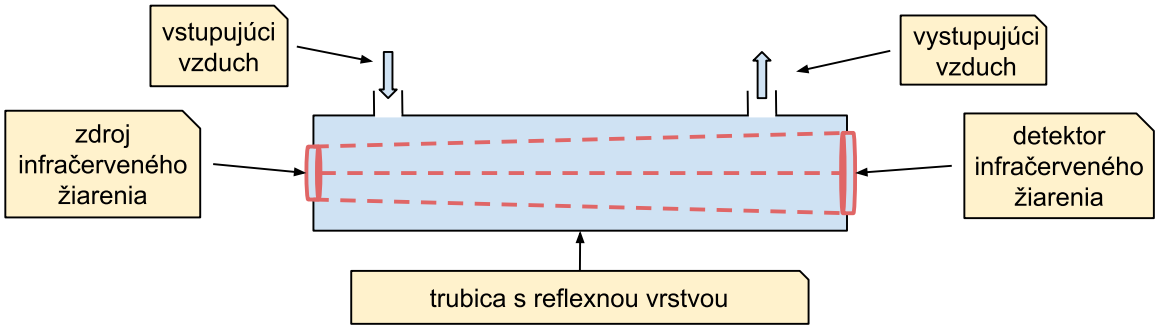
5.1 Prístroje pre meranie koncentrácie oxidu uhličitého

Meranie koncentrácie oxidu uhličitého môže prebiehať prístrojmi s rôznou technológiou a spôsobom merania. Meracie prístroje môžeme rozdeliť podľa charakteru ich použitia, a to na prenosné alebo stacionárne. Meranie okamžitej koncentrácie alebo kontinuálne meranie sú ďalšie dve kategórie spôsobu merania, do ktorých môžeme meracie prístroje kategorizovať. Najčastejšie používaným parametrom pre rozdelenie meracích prístrojov je samotný princíp, na ktorom senzory CO_2 pracujú. Dôležitým faktorom výberu senzorov je ich cena, životnosť a presnosť. Rozlišujeme nasledujúce senzory oxidu uhličitého [34]:

- optický (IR) senzor (NDIR – Non Dispesive Infra Red),
- elektrochemický senzor,
- elektroakustický senzor,
- polovodičový senzor,
- detekčná trubica.

Optický infračervený senzor pracuje na princípe merania absorpcie infračerveného (IR) spektra v molekulách CO_2. Meraním sa určuje zoslabenie IR žiarenia presne daných vl-

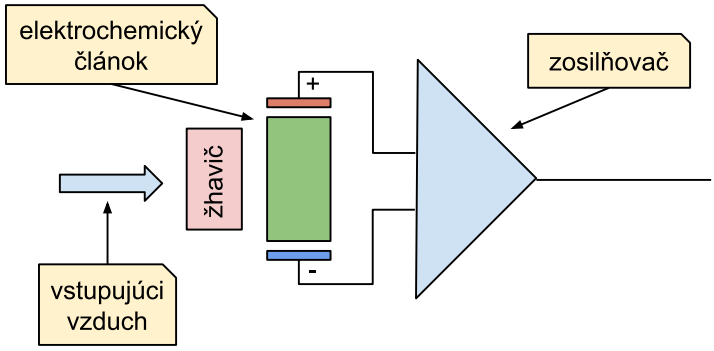
nových dĺžok (7.20 µm, 14.99 µm a 4.256 µm), ktoré slúžia pre výpočet koncentrácie CO_2 v meranom vzduchu. Senzor sa skladá zo zdroja infračerveného žiarenia, trubice s reflexnou vrstvou, infračerveného detektora, vstupu a výstupu pre meraný vzduch, v ktorom sa má zistiť koncentrácia. Jednotlivé časti senzora sú zobrazené na obrázku 5.1.



Obr. 5.1: Princíp fungovania infračerveného senzora na detekciu CO_2

Limitujúcim faktorom senzorov pracujúcich na optickom princípe je problematické meranie vyšších koncentrácií, pri ktorých môže dochádzať k pohlteniu všetkých sledovaných vlnových dĺžok. Spôsobený útlm vlnových dĺžok zníži kvalitu výstupného signálu a jeho celkovú presnosť. Výhodou týchto senzorov je však presné meranie nižších koncentrácii, dlhodobá stabilita merania a životnosť. Senzory umožňujú meranie od nulovej hodnoty a zvládnu aj meranie vysokých koncentrácii. Ich nevýhodou je vysoká cena [33].

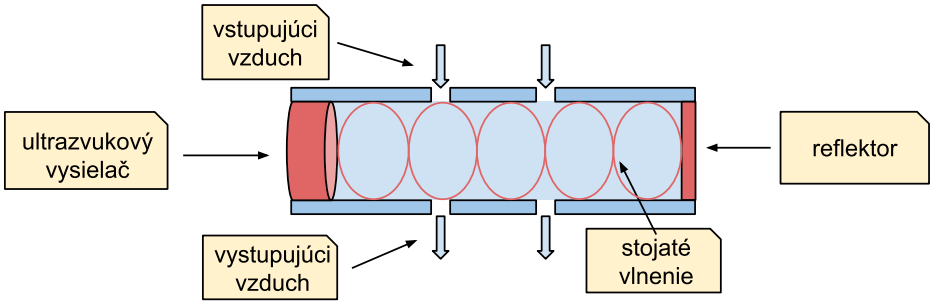
Elektrochemický senzor vytvára merateľný signál, ktorý je úmerný koncentrácii CO_2. Signál vzniká reakciou molekúl sledovaného plynu s elektrochemickým článkom obsahujúcim tuhý elektrolyt. Elektrolyt je potrebné zahriať na pracovnú teplotu pomocou žhaviča. Na elektródach tak dôjde k elektrochemickým reakciám, pri ktorých vznikajú voľné elektróny. Elektronika následne meria a zosilňuje vzniknutý signál, ktorý odpovedá nameranej koncentrácii CO_2 vo vzduchu. Senzor pozostáva zo žhaviča, elektrochemického článku, zosilňovača a meracej elektroniky [33]. Jednotlivé časti senzora sú zobrazené na obrázku 5.2.



Obr. 5.2: Princíp fungovania elektrochemického senzora na detekciu CO_2

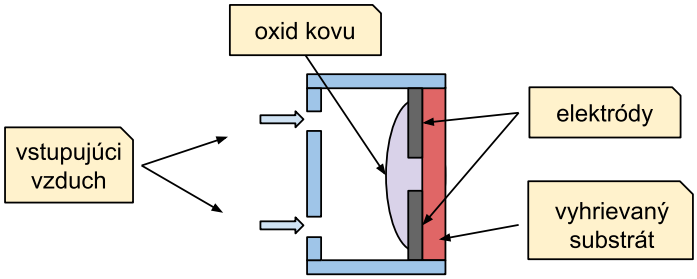
Výhodou elektrochemických senzorov je vysoká citlivosť na zmeny, vysoká rozlišovacia schopnosť CO_2 a vstavaná autokalibračná funkcia. Senzory umožňujú merať koncentráciu CO_2 približne len od 400 ppm, je vzhľadom ku koncentrácii 350–450 ppm CO_2 vo vonkajšom vzduchu postačujúce. Oproti infračerveným senzorom sú lacnejšie, ale ich životnosť je kratšia, čo čiastočne kompenzuje autokalibračná funkcia [33].

Elektroakustické senzory merajú zmenu kmitočtu ultrazvukových vĺn v mechanickom rezonátore, ktorá je úmerná koncentrácii CO_2. Senzor sa skladá z ultrazvukového vysielača a reflektora, ako je zobrazené na obrázku 5.3. Ultrazvukový vysielač vytvára stojaté vlnenie, ktoré vzniká odrazom vĺn od reflektora. Výhodou týchto senzorov je dlhá životnosť a stabilita meranej koncentrácie bez potreby kalibrácie [26].



Obr. 5.3: Princíp fungovania elektroakustického senzora na detekciu CO_2

Polovodičový senzor meria zmenu vodivosti polovodiča pomocou prítomnosti oxidu uhličitého a na základe zmeny detekuje koncentráciu CO_2. Senzor sa skladá z polovodiča, elektród a výhrevu pre substrát, pričom vhodnými materiálmi pre polovodiče sú oxidy kovov. Jednotlivé súčasti senzora sú zobrazené na obrázku 5.4. Nad povrchom polovodiča sa následne vytvori rovnovážny stav s molekulami kyslíka. Pri prítomnosti iného plynu sa naruší tento rovnovážny stav a zmení sa vodivosť medzi elektródami. Polovodičové senzory majú dlhú životnosť a nízku cenu. Medzi ich nevýhody patrí nízka rozlišovacia schopnosť CO_2 a časová nestabilita, ktorá sa dá vyriešiť dodatočnou elektronikou [33].



Obr. 5.4: Princíp fungovania polovodičového senzora na detekciu CO_2

Ďalšou možnosťou detekcie CO_2 je detekčná trubica, ktorá v sebe obsahuje chemické zlúčeniny a na svojom povrchu má zobrazenú stupnicu koncentrácie CO_2. Po privedení potrebného objemu vzduchu do detekčnej trubice nastane chemická reakcia a dôjde k jej zafarbeniu. Nameranú hodnotu je potom možné odčítať na stupnici. Špeciálnym prípadom je difúzna trubica, ktorá umožňuje zistiť priemernú koncentráciu CO_2 za určitý čas. Táto trubica sa ponechá požadovaný čas v meranom prostredí a priemerná koncentrácia sa vypočíta zo zafarbenia stupnice a času pôsobenia vzduchu na obsah trubice [33].

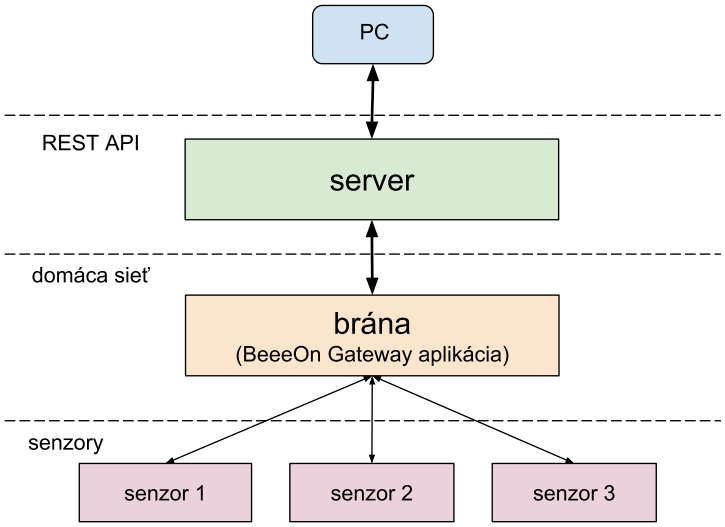
5.2 Snímače zmesi plynov

V rade prípadov môže nastať situácia, kedy je koncentrácia oxidu uhličitého v limitných hodnotách, ale ľudia sa sťažujú na kvalitu ovzdušia. Dôvodom nekvalitného vzduchu môžu byť aj rôzne zápachy, výpary, dym a ďalšie. Tieto látky sú nazývané ako tekové organické zlúčeniny alebo VOC (Volalite Organic Compounds). Snímače zmesi plynov, označované ako VOC, pracujú na princípe, kedy na rozžeravenej mriežke dochádza k ionizácií plynov a pracovná oblasť snímača mení svoj elektrický odpor. Tento elektrický odpor sa prevádza na výstupný signál. Snímač nemeria presnú hladinu CO_2, ale celkovú koncentráciu plynov v danom prostredí. Výstupná hodnota sa udáva v rozsahu 0 %–100 % a popisuje mieru znečistenia prostredia [34].

5.3 BeeeOn systém

Systém BeeeOn je open source projekt Fakulty informačných technológii Vysokého učenia technického v Brne. Výsledkom projektu boli dve hlavné aplikácie, a to serverová a klientská. Postupom času boli kladené nové a nové požiadavky na serverovú aplikáciu. Server na to architektonicky ani technologicky nebol pripravený, a preto bol v rámci projektu IoTCloud vytvorený nový server, ktorý spĺňa nové požiadavky a používa nové technológie. Projekt IoTCloud sa zameriava na analýzu nameraných dát a vytvorenie modelov pre rôzne odhady.

Projekt BeeeOn si kládol za cieľ vytvoriť systém, ktorým bude možné jednotným spôsobom ovládať a zbierať merané veličiny zo zariadení. Myšlienka jednotného ovládania vznikla z potreby spoločného ovládania zariadení od rôznych výrobcov, ktorí používajú rôzne protokoly, formáty dát ale i spôsoby ovládania. Zjednotením ovládania sa zobecnia základné operácie ako pridanie zariadenia do siete alebo odstránenie zariadenia zo siete, odoslanie hodnoty do zariadenia, či získanie nameraných hodnôt.



Obr. 5.5: Architektúra systému BeeeOn

Ako je vidno na obrázku 5.5, celý systém sa skladá z dvoch hlavných časti, a to zo serveru a brány. Na najspodnejšej vrstve sa nachádza brána. Bránou je typicky malý jednodoskový počítač, na ktorom je spustená BeeeOn Gateway aplikácia. Požiadavky na hardvér brány nie sú nijak špecifické a pre potreby projektu bol použitý open source hardvér s vývo-

jovou doskou A10-OLinuXino-LIME od spoločnosti Olimex s upraveným systémom, ktorý obsahoval len potrebné balíčky. Open source hardvér spolu s upraveným systémom a nainštalovanou BeeeOn Gateway aplikáciou sa nazýva BeeeOn Gateway. BeeeOn Gateway aplikáciu je možné preložiť a spustiť na ľubovoľnom systéme, ktorý podporuje potrebné knižnice pre preklad.

BeeeOn Gateway aplikácia je architektonicky rozdelená na samostatné komponenty. Medzi základné komponenty môžeme zaradiť manažéra zariadení, odosielateľa príkazov a distribútora. Odosielateľ príkazov zastrešuje správu príkazov, ktoré sa rozposielajú medzi jednotlivými komponentami v aplikácii. Príkaz je operácia alebo akcia, ktorou komponenty medzi sebou komunikujú. Správca zariadení poskytuje rozhranie pre pridanie nového protokolu (Z-Wave, IQRF, VPT) alebo množiny zariadení na určitom komunikačnom protokole (zariadenia komunikujúce pomocou Bluetooth od určitého výrobcu), s ktorými dokáže komunikovať. Pre komunikáciu s protokolmi sa najčastejšie využívajú špeciálne USB zariadenie, ktoré na jednej strane komunikuje zo senzorom v bezdrôtovej sieti a na druhej strane s BeeeOn Gateway aplikáciou. Správca zariadení musí implementovať operácie (pridanie, odstránenie zariadenia, nastavenie a získanie hodnoty) špecifické pre daný protokol tak, aby boli transparentné pre zvyšok systému. Musí tiež zaistiť prevod veličín do stanoveného formátu, prevod fyzikálnych veličín alebo zaokrúhlenie na požadovaný počet miest. Úlohou manažérov zariadení je teda prijaté dáta spracovať a previesť do formátu, ktorý je spoločný pre všetky protokoly. Manažéri zariadení posielajú spracované dáta zo senzorov do distribútora, ktorý ich exportuje prostredníctvom registrovaných exportérov. Exportné moduly môžu dáta ľubovoľne formátovať (JSON, CSV) a svoj výstup môžu posielať lokálne (linuxová rúra) alebo na server.

Hlavným účelom serveru je správa pripojených BeeeOn Gateway aplikácii, ukladanie nameraných dát a poskytnutie rozhrania pre vzdialené ovládanie. Dôležitú úlohu zohráva server aj pri získavaní nameraných dát zo senzorov, ktoré boli uložené na serveri. Server preto poskytuje REST API protokol pre správu brán a prístup k nameraným dátam.

V rámci projektu BeeeOn bol vytvorený vlastný senzor komunikujúci pomocou bezdrôtového protokolu, vyvinutým na Fakulte informačných technológii Vysokého učenia technického v Brne. Senzor je napájaný batériou a umožňuje merať teplotu a vlhkosť. K senzoru je možné pripojiť externé čidlo, ktoré slúži na meranie vonkajšej teploty. Pre podporu tohto senzora bola BeeeOn Gateway rozšírená o modul, ktorý v sebe obsahuje anténu pre komunikáciu so senzorom. Komunikácia so senzormi prebieha vo voľne dostupnom pásme 868–870 MHz s maximálnou prenosovou rýchlosťou 200 kb/s a maximálnym vysielacím výkonom 12,5 dBm. Modul komunikuje s bránou pomocou SPI rozhrania.

5.4 Výber senzorov

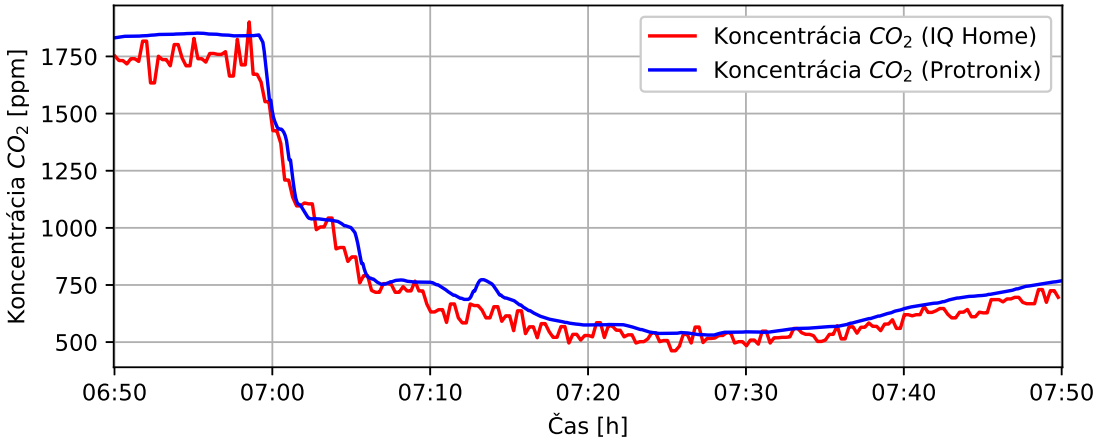
Systém BeeeOn podporuje dva druhy senzorov oxidu uhličitého, ktoré komunikujú bezdrôtovým protokolom IQRF. V tabuľke 5.1 je možné vidieť vlastnosti podporovaných senzorov od dvoch výrobcov, Protronix a IQ Home. Oba senzory umožňujú merať okrem koncentrácie CO_2 aj teplotu a vlhkosť. Protronix oproti IQ Home senzoru má menší rozsah merania, ale lepšiu presnosť a netrpí potrebou kalibrácie. Ak by sa senzor nekalibroval, dochádzalo by k postupnému nárastu minimálnej koncentrácie, ktorú senzor dokáže detekovať. Výhoda Protronix senzora je v napájaní externým zdrojom.

Zo senzorov uvedených v tabuľke 5.1 bol pre nasadenie preferovaný senzor od firmy Protronix, ktorý umožňuje napájanie externým zdrojom a namerané hodnoty koncentrácie CO_2 neoscilujú, ako v prípade senzoru od firmy IQ Home.

	Protronix NLII-CO2+RH+T-IQRF	**IQ Home SN-THC-02**
Podporované veličiny	CO_2, teplota, vlhkosť	CO2, teplota, vlhkosť
CO_2 rozsah [ppm]	0–2000	0–5000
CO_2 presnosť	$\pm$ 35 *ppm* $\pm$ 5% z hodnoty	$\pm$ 50 *ppm* $\pm$ 3% z hodnoty
Druh napájania	externým zdrojom	batériové napájanie

Tabuľka 5.1: Parametre senzorov od firiem Protronix a IQ Home [13, 25]

Priebeh koncentrácie CO_2 meranej oboma senzormi je zobrazený na obrázku 5.6. Pre orientačné meranie, kde nie je potrebná presnosť na desiatky ppm, je dostatočný aj senzor od firmy IQ Home, ktorý potrebuje pravidelnú kalibráciu.



Obr. 5.6: Nameraná koncentrácia CO_2 senzormi od firiem Protronix a IQ Home

Na presnú detekciu otvorenia okna bol použitý senzor Jablotron JA-81M, ktorý bol umiestnený na okennom ráme a detekoval otvorenie alebo zatvorenie okna. Pre získanie hodnôt zo senzora bolo potrebné USB zariadenie s názvom Turris Dongle, ktoré bolo pripojené k BeeeOn bráne. Toto zariadenie umožňovalo komunikáciu zo senzorom a zároveň s bránou. Senzor posielal hodnoty v prípade, že sa okno otvorilo alebo zatvorilo. Ak bolo okno otvorené, tak senzor navyše posielal informáciu o tomto stave každých 9 minút.

Pre meranie vnútorných a vonkajších podmienok bol použitý BeeeOn senzor. Umožňuje meranie teploty a vlhkosti, prípadne meranie teploty s externým čidlom. Senzor je napájaný batériami a pre meranie využíva čidlo SHT 21. Jedná sa o čidlo vlhkosti a teploty s presnosťou merania teploty $\pm 0.3°C$ a presnosťou merania vlhkosti $\pm 2\,\%$.

Kapitola 6

Zber a spracovanie dát

Cieľom bolo nazbierať dostatočné množstvo dát pre následné vytvorenie a otestovanie systému pre detekciu otvorenia okna a pre reguláciu výmeny vzduchu v miestnosti. Dáta boli namerané v niekoľkých miestnostiach, pričom dáta bolo potrebné spracovať do vhodnej podoby pre následnú analýzu.

6.1 Testovacie prostredia

Základom každej analýzy dát je dostatočný počet záznamov na vytvorenie a otestovanie modelu. Preto bolo potrebné získať dáta z niekoľkých testovacích prostredí, pričom hlavné testovacie prostredie slúžilo na prvotnú analýzu problému a jeho dlhodobé skúmanie. Ostatné testovacie prostredia slúžili na vyhodnotenie obecnosti vytvoreného modelu. Dáta boli zozbierané z kancelárskych a obytných priestorov.

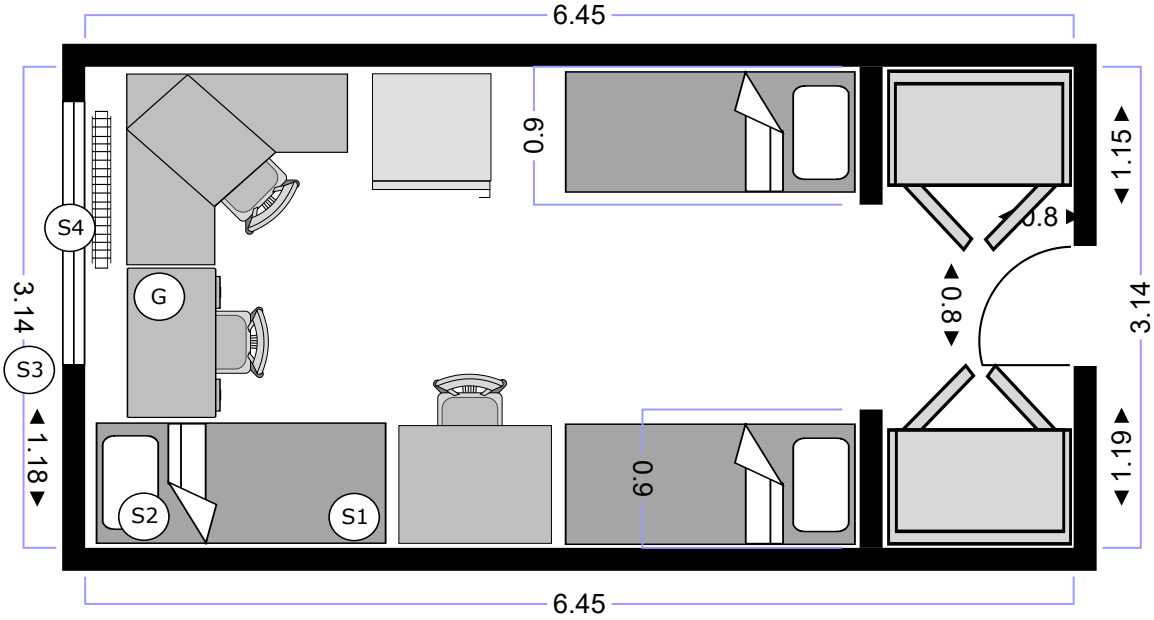
6.1.1 Hlavné testovacie prostredie

Ako hlavné testovacie prostredie bola zvolená obytná miestnosť, presnejšie sa jednalo o internátnu izbu pre tri osoby, ktorá sa nachádza v brnenskej mestskej časti Královo pole. Budova, kde sa nachádza sledovaná miestnosť, je nezrekonštruovaná, bez zateplenia s pôvodnými oknami. Izba je umiestnená na severovýchodnej strane a pred oknom je malý park. Meraná miestnosť je $6,45\,m$ dlhá, $3,14\,m$ široká, $2,70\,m$ vysoká a celkový objem miestnosti je približne $54,70\,m^3$. V miestnosti sa nachádza okno s dvomi krídlami. Šírka jedného krídla je $0,88\,m$ a jeho výška je $1,34\,m$. Pôdorys izby s približným rozložením nábytku a senzorov je zobrazený na obrázku 6.1.

Pre meranie vnútorných veličín boli použité tri senzory (S1, S2 a S4). Senzory S1 a S2 boli umiestnené približne 1,8 m nad podlahou na poličke. Senzor S4 bol umiestnený na ráme okna. Pre meranie vonkajších veličín bol použitý senzor S3, ktorý sa nachádzal na vonkajšom parapete.

Senzor S1 od firmy Protronix meral vnútornú teplotu, vlhkosť a koncentráciu CO_2. Jednalo sa o Protronix NLII-CO2+RH+T-IQRF. BeeeOn Senzory S2 a S3 umožňovali meranie teploty a vlhkosti. Senzor S4 od firmy Jablotron s označením JA-81M umožňoval detekciu otvorenia okna. Na stole bola umiestnená BeeeOn Gateway, v obrázku 6.1 označená písmenom G, ktorá umožňovala zber dát a ich odosielanie na server. Brána okrem jednodoskového počítača v sebe obsahuje aj senzor na meranie atmosferického tlaku.

Meranie vnútorných a vonkajších podmienok prebiehalo neustále. BeeeOn senzory posielali namerané údaje každých 15 sekúnd a Protronix senzor posielal údaje každých 5 sekúnd.



Obr. 6.1: Pôdorys hlavného testovacieho prostredia

Detektor otvorenia okna posielal údaje len pri otvorení alebo zatvorení okna. Okrem meraných veličín boli jednotlivé otvorenia okna anotované ďalšími informáciami, ako napríklad orientačná sila vetra, počet ľudí v miestnosti, aktivita ľudí v miestnosti, uhol otvorenia okna alebo, či slnko svieti na senzor, ktorý meria vonkajšie podmienky. Uhol otvorenia okna je dôležitým parametrom, od ktorého sa odvíja intenzita výmeny vzduchu. Počas vetrania sa poloha otvoreného krídla nemenila. Dlhodobé meranie prebiehalo od začiatku októbra 2018 do konca apríla 2019 (7 mesiacov).

6.1.2 Ostatné testovacie prostredia

Medzi ostatné testovacie prostredia patrili tri miestnosti, dve obytné miestnosti a jedna kancelárska miestnosť, ktoré sa nachádzali v Brne. Prostredia sa líšili svojou rozlohou, denným režimom a obsadenosťou.

Prvou testovacou miestnosťou je kancelárska miestnosť pre doktorandov na Fakulte informačných technológii. Miestnosť sa nachádza na prízemí zrekonštruovanej budovy v blízkosti križovatky a je situovaná na južnú stranu. Skúmaná miestnosť je $7\,m$ dlhá, $5,3\,m$ široká a $3,1\,m$ vysoká, jej približný objem je $115\,m^3$. V miestnosti sa nachádza niekoľko stolov, počítačov a príslušenstva k nim. Do miestnosti môže prúdiť čerstvý vzduch za pomoci štyroch okien, ale na vetranie sa používa len jedno, na ktorom bol osadený senzor na detekciu jeho otvorenia. Ďalej sa v miestnosti nachádzal senzor na meranie teploty, vlhkosti a koncentrácie CO_2, ktorý bol umiestnený približne v strede miestnosti vo výške $1,5\,m$ nad zemou. Počas dňa sa v miestnosti nachádzajú 3–4 ľudia, v noci sa v miestnosti nikto nenachádza. Meranie v tejto miestnosti prebiehalo od začiatku do konca apríla 2019 (1 mesiac).

Obytná miestnosť 1, situovaná na východnú stranu, sa nachádza v trojizbovom byte na prízemí. Miestnosť je $4,8\,m$ dlhá, $4\,m$ široká a $2,8\,m$ vysoká, jej objem je približne $53,8\,m^3$. V miestnosti sa nachádza pracovný stôl, odkladacie priestory a posteľ. Čerstvý vzduch môže do miestnosti prúdiť jedným oknom s výškou $1,4\,m$ a šírkou $0,5\,m$. Na ráme tohto okna

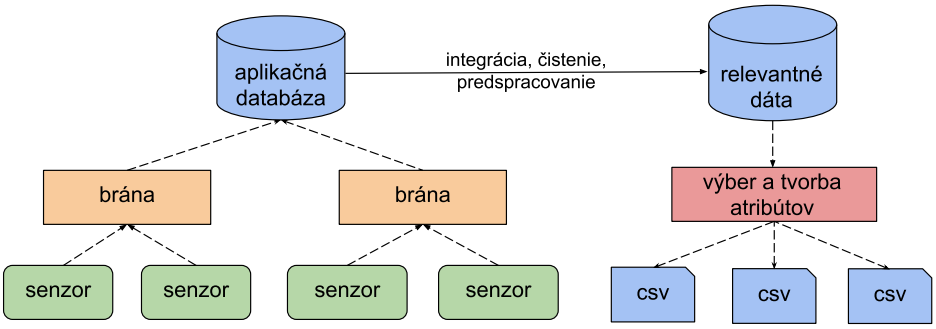
je umiestnený detektor jeho otvorenia a zároveň je v miestnosti senzor na meranie teploty, vlhkosti a koncentrácie CO_2, ktorý je umiestnený vo výške $1,7\,m$ nad zemou a asi $0,2\,m$ od steny. Senzor sa nachádza na kraji izby, avšak nie v jej rohu. Počas dňa a noci sa v miestnosti nachádza 1–2 ľudia. Meranie v tejto miestnosti prebiehalo od polovice decembra 2018 do konca apríla 2019 (4,5 mesiaca).

Obytná miestnosť 2 je izba v trojizbovom byte na druhom poschodí, situovaná na severnú stranu. Miestnosť je $7\,m$ dlhá, $5\,m$ široká a $2,5\,m$ vysoká s približným objemom $87,5\,m^3$. V priestore sa nachádzajú odkladacie priestory, pracovný stôl a posteľ. Čerstvý vzduch sa do miestnosti privádza pomocou okna so šírkou $0,7\,m$ a výškou $1,6\,m$. Na detekciu otvorenia okna sa používa senzor umiestnený na okennom ráme. Teplota, vlhkosť a koncentrácia CO_2 je meraná pomocou senzoru umiestneného naproti oknu, na druhej strane miestnosti, vo výške približne $1,8\,m$ na zemou. Počas dňa a noci sa v miestnosti nachádza jeden človek. Meranie v tomto priestore prebiehalo od začiatku do konca apríla 2019 (1 mesiac).

6.2 Spracovanie nameraných hodnôt

Pre potreby analýzy dát boli namerané hodnoty zo senzorov zbierané BeeeOn Gateway bránou, ktorá okrem ich zberu zabezpečovala aj ich odosielanie na server. Server namerané hodnoty ukladal do databáze a umožňoval k nim pristupovať pomocou REST API.

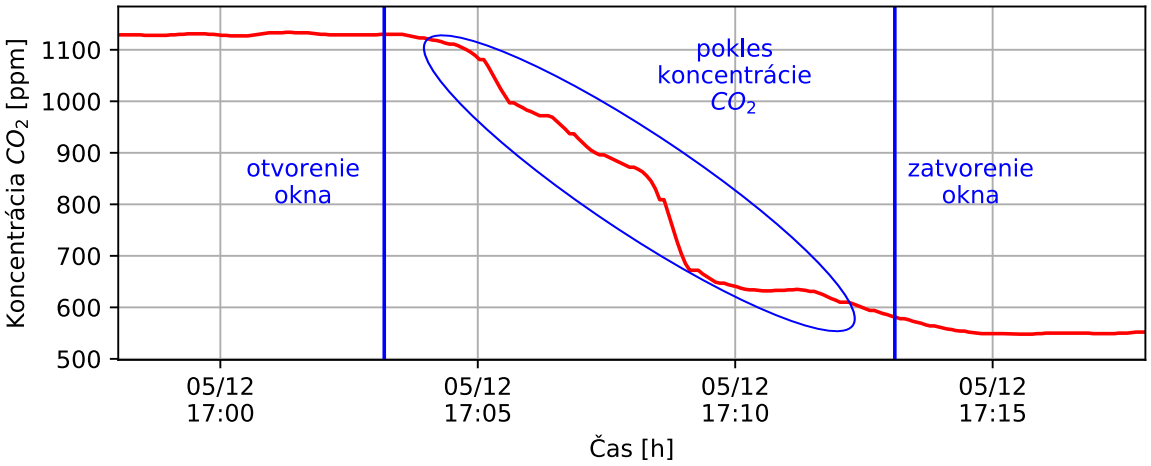
Počas zberu nameraných dát sa občas vyskytli problémy so senzormi v podobe krátkych výpadkov, kedy senzory nemerali žiadne dáta, alebo posielali nevalidné dáta. Okrem spomenutých problémov bolo potrebné vyriešiť problém, ako získať požadovaný interval hodnôt zo serveru, prípadne nejakú konkrétnu hodnotu v určitom čase s presnosťou na jednu sekundu. Dôvodom spomenutého problému bolo to, že senzory neposielajú namerané hodnoty každú sekundu, ale len niekoľkokrát za minútu. Riešením bolo jednoduché dopočítanie chýbajúcich hodnôt, pričom závislosť medzi dvomi hodnotami bola zjednodušená na lineárnu. Ďalej bolo potrebné navrhnúť spôsob, akým sa budú dáta pre analýzu získavať. Dotazovanie sa na server pri každej práci s určitým intervalom nameraných hodnôt s presnosťou jednej sekundy je časovo a výpočetne náročné. Lepšou voľbou je stiahnutie nameraných dát, ich predspracovanie a uloženie do lokálnej databáze, ktorá obsahuje údaje vzorkované každú sekundu (obrázok 6.2). Potom je možné na základe dát rýchlo vypočítať atribúty a výstup uložiť do súboru v CSV formáte, ako je zobrazené na obrázku 6.2. Tento súbor je potom použitý ako vstup pre učenie alebo testovanie modelov v programe RapidMiner.



Obr. 6.2: Schéma predspracovania nameraných hodnôt zo senzorov

6.3 Tvorba a výber atribútov

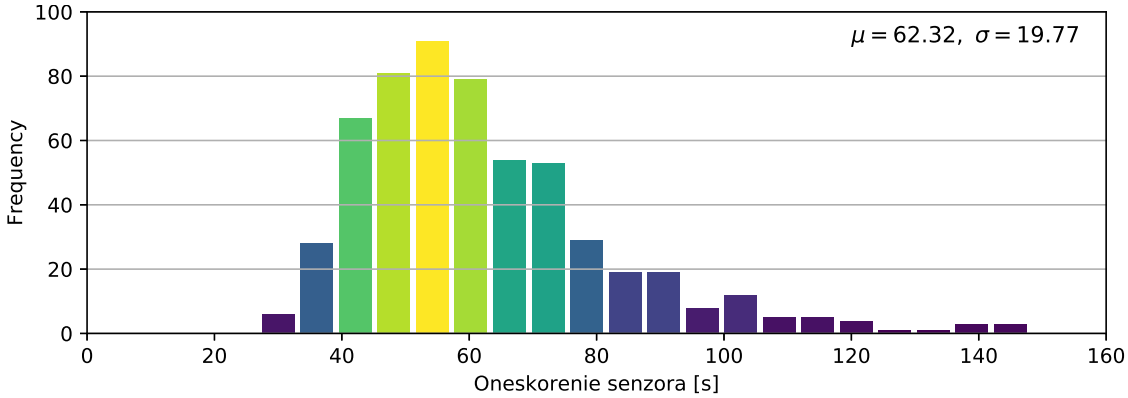
Úspešnosť klasifikácie je závislá nie len od dostatočného počtu dát ale aj od atribútov, ktoré musia dobre popisovať klasifikovanú udalosť a jej okolie. V prípade detekcie otvorenia okna je potrebné vytvoriť atribúty, ktoré umožnia vyjadriť zmenu koncentrácie CO_2. Typický priebeh koncentrácie počas vetrania (otvorenie, zatvorenie okna) je zobrazený na obrázku 6.3.



Obr. 6.3: Typický priebeh koncentrácie CO_2 počas vetrania

Ako je vidno na obrázku 6.3, najjednoduchšími atribútmi by boli okamžité hodnoty koncentrácie CO_2. Okamžité hodnoty majú nevýhodu v tom, že nezohľadňujú veľkosť zmeny koncentrácie počas vetrania. Riešením je vyjadrenie poklesu pomocou zmeny koncentrácie medzi určitými hodnotami v čase. Takýto pokles je možné vyjadriť pomocou diferencii, tempa rastu alebo pomocou ďalších funkcii.

Detekciu otvorenia okna nie je možné určiť hneď po otvorení okna ale až s určitým oneskorením. Je to dané tým, že koncentrácia CO_2 začne postupne klesať až po určitom čase, ktorý je daný princípom merania a umiestnením senzora. Čím je senzor umiestnený ďalej od okna, tým trvá dlhšie, než sa čerstvý vzduch dostane k senzoru. Pri analýze hlavnej trénovacej sady (568 otvorení okna) bolo zistené, že začiatok poklesu koncentrácie CO_2 začína minimálne 25 sekúnd po otvorení okna, ako je zobrazené v histograme na obrázku 6.4.



Obr. 6.4: Histogram začiatku poklesu koncentrácie CO_2

Stredná hodnota oneskorenia poklesu koncentrácie CO_2 je 62,3 sekúnd so smerodatnou odchýlkou 19,8 sekúnd.

Záznamy pre trénovanie a testovanie sa vytvárajú tak, že je určená veľkosť posuvného okna, z ktorého sa potom vytvoria atribúty. Posuvné okno obsahuje historické hodnoty, pričom určitá časť týchto historických hodnôt je daná oneskorením, ktoré je potrebné pre detekciu otvorenia okna.

Každý záznam, či už pre učenie alebo testovanie, môže pozostávať z nasledujúcich typov atribútov:

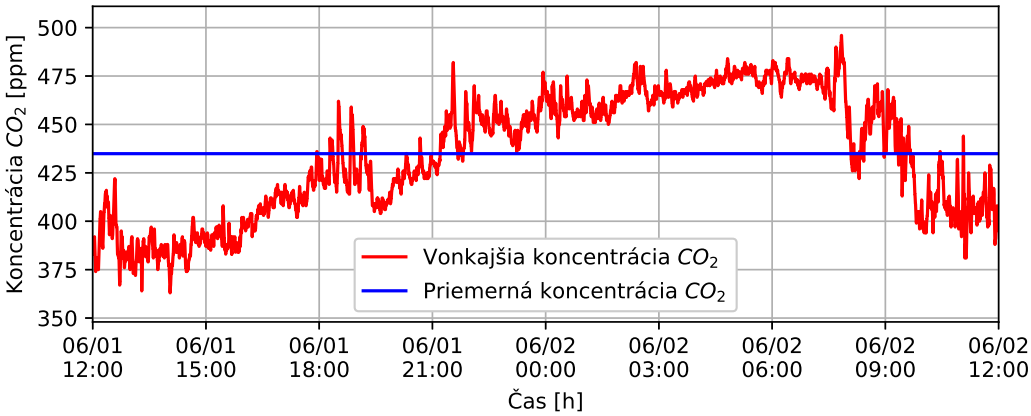
differencie – rozdiel medzi dvomi hodnotami v čase, pričom je možné rozdiel vypočítať medzi susednými hodnotami alebo medzi hodnotami a jednou konkrétnou hodnotou,

rozdiel medzi skutočnou a linearizovanou hodnotou – linearizácia určitého časového úseku a následný rozdiel medzi zlinearizovanou a skutočnou hodnotou v danom čase,

tempo rastu – percentuálna zmena hodnôt danej veličiny za určitý časový interval,

exponenciálna regresia – hodnota funkcie predpísaného tvaru (viď rovnica (2.9)), ktorá vyžaduje parameter λ a namerané hodnoty,

rozdiel medzi hodnotou nameranou vo vnútri a vonku – v prípade, že sa jednalo o rozdiel koncentrácií CO_2, tak bola vonkajšia hodnota koncentrácie stanovená na základe priemeru hodnôt nameraných za 24 hodín na 434 ppm, jej priebeh je zobrazený na obrázku 6.5.



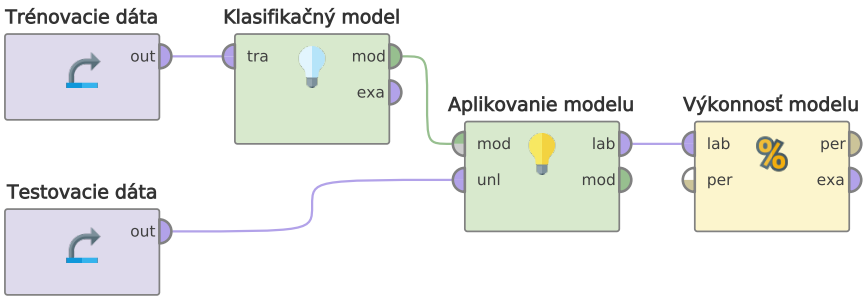
Obr. 6.5: Koncentrácia CO_2 vo vonkajšom vzduchu, priemerne 434 ppm

Kapitola 7

Detekcia otvorenia okna

Cieľom úlohy bolo vytvoriť detektor, ktorý na základe zmeny koncentrácie CO_2 dokáže detekovať otvorenie okna. Pred samotným vytvorením detektora bolo potrebné vybrať vhodné atribúty, ktoré umožnia popísať pokles koncentrácie. Detektory popísané nižšie využívajú na zhotovenie modelu atribúty vytvorené z nameraných hodnôt koncentrácie CO_2, vnútornej a vonkajšej teploty a vlhkosti.

Pre vytvorenie detektora na detekciu otvorenia okna bol použitý program RapidMiner, kde bola vytvorená bloková schéma procesu, ktorá je zobrazená na obrázku 7.1. Jednotlivé bloky reprezentujú kroky, potrebné pre vytvorenie detektora.



Obr. 7.1: Bloková schéma detektoru otvorenia okna v programe RapidMiner

Vstupom procesu sú trénovacie dáta, ktoré obsahujú záznamy pre učenie, a testovacie dáta so záznamami pre testovanie. Každý záznam pozostáva z identifikátora (*id*), označenia, do ktorej triedy záznam patrí (*label*), a príslušných atribútov. Trénovacie dáta slúžia ako vstup do klasifikačnej metódy, ktorou bolo hlboké učenie, metóda podporných vektorov (SVM), rozhodovací strom, náhodný les, Bayesovská klasifikácia alebo neurónová sieť. Výsledkom klasifikačnej metódy je detekčný model, na ktorý sú následne aplikované testovacie dáta. Na konci procesu sa vyhodnotí výkonnosť modelu, v podobe úspešnosti klasifikácie dát do správnych tried.

Vyhodnotenie úspešnosti sa vykoná pomocou matice zámen, ktorá ukazuje:

- true positive (TP) – koľko výsledkov klasifikácie bolo skutočne pozitívnych, teda koľko otvorení okien bolo správne klasifikovaných ako otvorenie okna,
- false negative (FN) – koľko výsledkov klasifikácie bolo falošne negatívnych, teda koľko otvorení okien bolo chybne klasifikovaných ako žiadna akcia,

- true negative (TN) – koľko výsledkov klasifikácie bolo skutočne negatívnych, teda koľko žiadnych akcii bolo správne klasifikovaných ako žiadna akcia,
- false positive (FP) – koľko výsledkov klasifikácie bolo falošne pozitívnych, teda koľko žiadnych akcii bolo chybne klasifikovaných ako otvorenie okna.

Celková úspešnosť detektora (*klasický prístup*), vypočítaná ako podiel správne klasifikovaných záznamov a všetkých záznamov, nie je vhodná pre vyjadrenie úspešnosti

$$klasicka_uspesnost = \frac{TP + TN}{TP + FN + TN + FP}. \tag{7.1}$$

Dôvodom je, že sa pracuje s časovými radami, kedy môže byť otvorenie okna detekované s určitým predstihom, oneskorením, alebo niekoľkokrát za sebou v určitom okolí skutočného otvorenia okna. Riešením tohto problému môže byť upravený výpočet úspešnosti (*upravený prístup*). Úprava spočíva v zavedení tolerancie, kedy budú ignorované falošné hlásenia otvorenia okna v okolí jeho skutočného otvorenia. Falošné hlásenia (false positives) v rámci tolerancie sú označené ako true negatives. Celková úspešnosť detektoru sa opäť vypočíta ako podiel správne klasifikovaných záznamov a všetkých záznamov (viď rovnica (7.1)).

I po zavedení tolerancie bolo detekované veľké množstvo chybných hlásení otvorenia okna. Problém spočíval v nekvalitnej trénovacej sade. Trénovacia sada musí obsahovať vyvážený počet záznamov, pričom záznamy, označujúce, že nedošlo k otvoreniu okna, sú generované automaticky. Záznamy, kedy nedošlo k otvoreniu okna, sú reprezentované atribútmi, vypočítanými na základe historických hodnôt, ktoré boli získane niekoľko desiatok sekúnd pred otvorením okna. Preto nie je možné zachytiť rôzne typy priebehov koncentrácie CO_2, ktoré sú spôsobené ďalšími aktivitami ako spánok, odchod z miestnosti alebo otvorenie dverí. Problém je možné eliminovať náhradou napevno vygenerovaných záznamov nesprávne detekovanými otvoreniami okna. Vzhľadom na to, že trénovacia sada musí byť vyvážená, je možné vyššie uvedeným spôsobom nahradiť len obmedzený počet záznamov, a preto nie je možné pridať všetky nesprávne detekované otvorenia okna. Postupnými iteráciami, teda nahradzovaním malého počtu napevno vygenerovaných záznamov nesprávne detekovanými otvoreniami okna, je možné, i v prípade malých trénovacích sad, veľmi rýchlo zvýšiť úspešnosť detektorov.

Detektor otvorenia okna bol vytvorený na základe troch sad atribútov. Cieľom bolo zistiť, či zahrnutím ďalších veličín dôjde k zvýšeniu úspešnosti detektora. Prvá sada atribútov bola vytvorená pomocou nameraných hodnôt koncentrácie CO_2, druhá sada naviac uvažovala namerané hodnoty vnútornej teploty a vlhkosti a tretia sada zahŕňala hodnoty vyššie spomenutých veličín a namerané hodnoty vonkajšej teploty a vlhkosti.

Pre fázu trénovania detektorov boli použité dáta z hlavného testovacieho prostredia, namerané za obdobie od začiatku októbra 2018 do začiatku februára 2019 (4 mesiace). Pre vyhodnotenie obecnosti detektorov boli použité štyri testovacie sady, ktoré boli získané z kancelárskej miestnosti a troch obytných priestorov (viď podsekcia 6.1.2). Každá testovacia sada neobsahovala záznamy vzorkované každú sekundu. Dôvodom bolo, že pre detekciu zmeny koncentrácie CO_2, v prípade reálneho nasadenia, postačí posielať hodnoty zo senzorov každých 30 sekúnd.

Vyhodnotenie úspešnosti klasifikátorov bolo vykonané na základe dvoch prístupov. Prvým spôsobom bolo klasické vyhodnotenie úspešnosti (viď rovnica (7.1)). Pre vyhodnotenie úspešnosti bolo však vzhľadom k cieľovej doméne vhodnejšie použiť upravenú úspešnosť (*upravený prístup*). Tolerancia bola nastavená tak, že falošné hlásenia otvorenia okna 2 minúty pred a 3 minúty po jeho otvorení boli ignorované.

Tabuľky 7.1–7.14 reprezentují dosiahnuté úspešnosti jednotlivých klasifikátorov, vrátane TP, FN, TN, FP. Medzi uvažované klasifikátory patria hlboké učenie, SVM, rozhodovací strom, náhodný les, Bayesovský klasifikácia a neurónová sieť.

7.1 Detekcia na základe koncentrácie oxidu uhličitého

Trénovacia sada obsahovala 1 140 záznamov a testovacia sada zahŕňala 237 974 záznamov. Jednotlivé záznamy obsahovali 443 atribútov, pričom v trénovacej sade bol každý typ atribútu definovaný na základe hodnôt z posuvného okna o veľkosti 23 minút, z čoho 3 minúty tvorilo oneskorenie. Tabuľka 7.1 reprezentuje dosiahnuté úspešnosti klasifikátorov v prípade detekcie otvorenia okna na základe koncentrácie oxidu uhličitého. Dôvodom vysokej úspešnosti klasifikátorov je veľký počet true negatives. Počet false negatives je oproti počtu false positives zanedbateľný. Vytvorený detektor je schopný dobre detekovať otvorenie okna, ale vo veľa prípadoch dochádza k falošnému detekovaniu otvorenia okna. Najvyššej upravenej úspešnosti (98,91 %) a zároveň najmenšieho počtu false positives dosiahol detektor vytvorený pomocou SVM klasifikátora.

	Prístup	**Úspešnosť [%]**	**True positive**	**False negative**	**True negative**	**False positive**
Hlboké učenie	klasický	89,90	333	5	213 601	24 035
	upravený	91,01	338	0	216 252	21 384
SVM	klasický	97,99	329	9	232 855	4 781
	upravený	98,91	336	2	235 040	2 596
Rozhodovací strom	klasický	96,28	337	1	228 778	8 858
	upravený	97,47	338	0	231 621	6 015
Náhodný les	klasický	96,37	336	2	228 991	8 645
	upravený	97,63	338	0	232 005	5 631
Bayesovská klasifikácia	klasický	97,88	336	2	215 804	4 669
	upravený	98,62	338	0	217 433	3 040
Neurónová sieť	klasický	97,25	337	1	231 090	6 546
	upravený	98,39	338	0	233 797	3 839

Tabuľka 7.1: Dosiahnuté úspešnosti klasifikátorov pri počiatočnom vyhodnotení v hlavnom testovacom prostredí (uvažovaná len koncentrácia CO_2)

S cieľom znížiť počet false positives bol vykonaný iteratívny proces učenia detektora. Po vykonaní 58. iterácie boli odstránené všetky false positives (viď tabuľka 7.2), pričom v každej iterácii bolo v trénovacej sade nahradených 10 záznamov. Za cenu nulového počtu false positives mierne vzrástol počet false negatives. Najvyššiu upravenú úspešnosť (100,00 %) a najmenší počet false negatives dosiahol detektor vytvorený na základe hlbokého učenia. Ostatné klasifikátory dosiahli vyššieho počtu false negatives, ich počet môže byť ďalej ovplyvnený nastavením parametrov jednotlivých klasifikátorov.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	99,37	335	3	236 133	1 503
	upravený	100,00	338	0	237 636	0
SVM	klasický	99,24	320	18	235 842	1 794
	upravený	100,00	328	10	237 636	0
Rozhodovací strom	klasický	99,50	313	25	236 475	1 161
	upravený	99,99	321	17	237 636	0
Náhodný les	klasický	99,32	329	9	236 030	1 606
	upravený	100,00	332	6	237 636	0
Bayesovská klasifikácia	klasický	99,47	329	9	233 566	1 233
	upravený	100,00	332	6	234 799	0
Neurónová sieť	klasický	99,28	331	7	235 940	1 696
	upravený	100,00	335	3	237 636	0

Tabuľka 7.2: Dosiahnuté úspešnosti klasifikátorov po 58. iterácii v hlavnom testovacom prostredí (uvažovaná len koncentrácia CO_2)

Naučený detektor otvorenia okna bol otestovaný na dátach, ktoré boli namerané v hlavnej testovacej miestnosti (viď podsekcia 6.1.1) a zároveň neboli použité pre iteratívne trénovanie. Testovacia sada obsahovala 189 526 záznamov, ktoré boli vytvorené na základe dát nameraných od začiatku februára 2019 do konca apríla 2019. Tabuľka 7.3 reprezentuje najlepší výsledok, dosiahnutý za použitia náhodného lesa (99,99 %). Výsledky dosiahnuté všetkými klasifikátormi sú uvedené v tabuľke B.1, ktorá je dostupná v prílohe B. S využitým náhodného lesa bolo detekované 8 false negatives a 5 false positives. Dôvodom nesprávnej klasifikácie bolo príliš krátke vetranie alebo otvorenie okna pri nízkej počiatočnej koncentrácie CO_2, kedy je priebeh veľmi podobný priebehu pri odchode osoby z miestnosti.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Náhodný les	klasický	99,13	338	13	187 434	1 642
	upravený	99,99	343	8	189 071	5

Tabuľka 7.3: Úspešnosť náhodného lesa pri obecnom otestovaní v hlavnom testovacom prostredí (uvažovaná len koncentrácia CO_2)

Ďalej boli pre otestovanie klasifikátorov použité dáta, získané z kancelárskej miestnosti počas apríla 2019 (viď podsekcia 6.1.2). Testovacia sada zahŕňala 41 813 záznamov. Najlepšie výsledky boli získané pomocou náhodného lesa (99,98 %, viď tabuľka 7.4). Úspešnosti všetkých klasifikátorov sú obsiahnuté v tabuľke B.2, ktorá je dostupná v prílohe B. Neboli detekované žiadne falošné hlásenia otvorenia okna, ale zároveň nebolo správne detekované žiadne z 10 otvorení okna. Príčinou môže byť príliš mierny pokles koncentrácie CO_2 po otvorení okna alebo vetranie menšou plochou otvárateľnej časti okna ako v hlavnom testovacom prostredí.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Náhodný les	klasický	99,98	0	10	41 793	0
	upravený	99,98	0	10	41 793	0

Tabuľka 7.4: Úspešnosť náhodného lesa pri obecnom otestovaní v kancelárskej miestnosti (uvažovaná len koncentrácia CO_2)

Testovacia sada, obsahujúca 283 508 záznamov, bola vytvorená na základe dát, získaných počas merania v obytnej miestnosti 1, a to od konca novembra 2018 do konca apríla 2019. Z vybraných klasifikátorov dosiahla najlepších výsledkov neurónová sieť (99,96 %, viď tabuľka 7.5), pričom výsledky všetkých klasifikátorov sú dostupné v tabuľke B.3, ktorá je dostupná v prílohe B. Celkom bolo falošne detekovaných 70 otvorení okna, pričom len 35 otvorení okna nebolo detekovaných vôbec. Dôvodom nesprávnej detekcie bolo občasné vetranie v iných miestnostiach pri otvorených dverách, ale aj menšia plocha otvárateľnej časti okna ako v hlavnom testovacom prostredí.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Neurónová sieť	klasický	99,36	258	183	281 300	1 641
	upravený	99,96	406	35	282 871	70

Tabuľka 7.5: Úspešnosť neurónovej siete pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná len koncentrácia CO_2)

Posledným testovacím prostredím bola obytná miestnosť 2, kde boli dáta zhromažďované v priebehu apríla 2019. Testovacia sada sa skladala z 59 614 záznamov. Detekcia otvorenia okna bola najúspešnejšia v prípade použitia neurónovej siete (99,84 %, viď tabuľka 7.6), pričom úspešnosti všetkých klasifikátorov sú uvedené v tabuľke B.4, ktorá je dostupná v prílohe B. Počet false negatives bol 19 a počet false positives bol 78. Nesprávna klasifikácia bola spôsobená pomalým poklesom koncentrácie CO_2, väčším objemom miestnosti a menšou plochou otvárateľnej časti okna ako v hlavnom testovacom prostredí. Pomalý pokles môže byť tiež spôsobený aj umiestnením senzora v opačnej časti miestnosti ako je okno.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Neurónová sieť	klasický	99,78	2	28	59 467	103
	upravený	99,84	11	19	59 492	78

Tabuľka 7.6: Úspešnosť neurónovej siete pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná len koncentrácia CO_2)

7.2 Detekcia na základe koncentrácie oxidu uhličitého, vnútornej teploty a vlhkosti

Trénovacia sada pozostávala z 1 140 záznamov, testovacia sada z 238 157 záznamov. V oboch sadách záznamy zahŕňali 1 327 atribútov, pričom veľkosť posuvného okna bola opäť 23 minút, z toho 3 minúty predstavovali oneskorenie. Úspešnosti klasifikátorov pri detekcii otvorenia okna, založené na koncentrácii CO_2, vnútornej teplote a vlhkosti sú ukázané v tabuľke 7.7. Vysoká úspešnosť klasifikátorov bola opäť spôsobená značným počtom true negatives. Bolo však potrebné eliminovať vysoký počet false positives, počet false negatives bol nízky. Detektor umožňuje dobre detekovať otvorenie okna, ale zároveň mnohokrát nesprávne detekuje jeho otvorenie. Neurónová sieť, použitá pre vytvorenie detektoru, dosiahla najlepšej upravenej úspešnosti (99,86 %) a najmenej false positives.

	Prístup	**Úspešnosť [%]**	**True positive**	**False negative**	**True negative**	**False positive**
Hlboké učenie	klasický	90,45	338	0	214 918	22 718
	upravený	91,78	338	0	218 070	19 566
SVM	klasický	97,73	332	6	232 230	5 406
	upravený	98,86	337	1	234 921	2 715
Rozhodovací strom	klasický	96,51	338	0	229 322	8 314
	upravený	97,88	338	0	232 598	5 038
Náhodný les	klasický	96,92	338	0	230 305	7 331
	upravený	98,18	338	0	233 310	4 326
Bayesovská klasifikácia	klasický	98,43	337	1	216 577	3 451
	upravený	99,17	338	0	218 205	1 823
Neurónová sieť	klasický	98,91	334	4	235 055	2 581
	upravený	99,86	336	0	237 311	325

Tabuľka 7.7: Dosiahnuté úspešnosti klasifikátorov pri počiatočnom vyhodnotení v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť)

Pomocou iteratívneho procesu učenia detektora boli po 58 iteráciách odstránené všetky false positives. V rámci každej iterácii bolo nahradených 10 záznamov. Na konci iteratívneho procesu bol detekovaný väčší počet false negatives. Detektor vytvorený pomocou hlbokého učenia dosiahol najlepších výsledkov (100,00 %). Ostatné klasifikátory dosahovali tiež veľmi vysokých úspešností, ale počet false negatives bol mierne vyšší (viď tabuľka 7.8). Zníženie počtu false negatives pri ostatných klasifikátoroch by bolo možné dosiahnuť nastavením ich parametrov.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	99,24	334	4	235 832	1 804
	upravený	100,00	337	1	237 636	0
SVM	klasický	99,28	324	14	235 934	1 702
	upravený	100,00	330	8	237 636	0
Rozhodovací strom	klasický	99,31	315	23	236 028	1 608
	upravený	99,99	321	17	237 636	0
Náhodný les	klasický	99,37	331	7	236 143	1 493
	upravený	100,00	335	3	237 636	0
Bayesovská klasifikácia	klasický	99,46	333	5	235 355	1 265
	upravený	100,00	335	3	236 616	0
Neurónová sieť	klasický	99,19	333	5	235 712	1 924
	upravený	100,00	335	3	237 636	0

Tabuľka 7.8: Dosiahnuté úspešnosti klasifikátorov po 58. iterácii v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť)

Po natrénovaní bol detektor otvorenia okna testovaný na dátach, získaných z hlavnej testovacej miestnosti (viď podsekcia 6.1.1), ktoré neboli využité pre iteratívne trénovanie. Testovacia sada zahŕňala 189 526 záznamov, ktoré boli definované pomocou dát nameraných v období od začiatku februára 2019 do konca apríla 2019. Najlepšie výsledky dosiahol detektor natrénovaný pomocou náhodného lesa (100,00 %, viď tabuľka 7.9). Úspešnosti všetkých uvažovaných klasifikátorov sú prezentované v tabuľke B.5, ktorá je dostupná v prílohe B. Počet nesprávnych detekcii otvorenia okna bol 4, počet nedetekovaných otvorení okna bol tiež 4.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Náhodný les	klasický	99,18	342	9	187 539	1 537
	upravený	100,00	347	4	189 072	4

Tabuľka 7.9: Úspešnosť náhodného lesa pri obecnom otestovaní v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť)

Ďalšie testovacie prostredie predstavovala kancelárska miestnosť, kde boli dáta zbierané počas apríla 2019. Testovacia sada obsahovala celkom 41 813 záznamov. Spomedzi všetkých klasifikátorov bol najúspešnejší rozhodovací strom (99,98 %, viď tabuľka 7.10). Výsledky všetkých klasifikátorov sú ukázané v tabuľke B.6, ktorá je dostupná v prílohe B. Detektor nedetekoval chybne žiadne otvorenia okna, ale počet nedetekovaných otvorení okna bol 9.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Rozhodovací strom	klasický	99,97	0	10	41 789	4
	upravený	99,98	1	9	41 793	0

Tabuľka 7.10: Úspešnosť rozhodovacieho stromu pri obecnom otestovaní v kancelárskom prostredí (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť)

Testovacia sada s 216 594 záznamami bola vytvorená pomocou dát nameraných v období od konca novembra 2018 do konca apríla 2019. Meranie prebiehalo v obytnej miestnosti 1 (viď podsekcia 6.1.2). Najlepšie výsledky, ktoré sú prezentované v tabuľke 7.11, vykazoval náhodný les (99,96 %). Úspešnosti všetkých klasifikátorov sú dostupné v tabuľke B.7, ktorá je dostupná v prílohe B. Počet false positives bol 27 a počet false negatives bol 49.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Náhodný les	klasický	99,57	248	107	215 193	834
	upravený	99,96	306	49	216 000	27

Tabuľka 7.11: Úspešnosť náhodného lesa pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť)

Na základe dát, získaných z obytnej miestnosti 2 počas apríla 2019, bola vytvorená testovacia sada, ktorá zahrňovala 59 614 záznamov. Detektor otvorenia okna dosahoval najvyššiu úspešnosť v prípade natrénovania pomocou neurónovej siete (99,95 %, viď tabuľka 7.12). Všetky klasifikátory s ich dosiahnutými úspešnosťami sú zobrazené v tabuľke B.8, ktorá je dostupná v prílohe B. Boli detekované len 2 chybné otvorenia okna z 28 a otvorenie okna bolo detekované v 4 prípadoch.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Neurónová sieť	klasický	99,94	0	30	59 564	6
	upravený	99,95	2	28	59 566	4

Tabuľka 7.12: Úspešnosť neurónovej siete pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť)

Faktory ovplyvňujúce počet false negatives a false positives sú rovnaké ako v predchádzajúcom prípade. Medzi najčastejšie príčiny patrí príliš malý pokles koncentrácie CO_2 po otvorení okna, spôsobený väčším objemom miestnosti alebo menšou plochou otvárateľnej časti okna ako v hlavnom testovacom prostredí. Dosiahnuté výsledky nedokázali zlepšenie úspešnosti detekcie otvorenia okna, pokiaľ sa berie do úvahy i vnútorná teplota a vlhkosť.

7.3 Detekcia na základe koncentrácie oxidu uhličitého, vnútornej a vonkajšej teploty a vlhkosti

Celkom 1 096 záznamov bolo použitých pre vytvorenie trénovacej sady, testovacia sada obsahovala 214 855 záznamov. Záznamy sa skladali z 1 773 atribútov a boli vypočítané v rámci posuvného okna o veľkosti 23 minút, z čoho 3 minúty tvorili oneskorenie. V prípade detekcie otvorenia okna podľa koncentrácie CO_2, vnútornej a vonkajšej teploty a vlhkosti sú úspešnosti dosiahnuté jednotlivými klasifikátormi zobrazené v tabuľke 7.13. Úspešnosť klasifikátorov bola vysoká hlavne preto, že bol detekovaný veľký počet false negatives. Otvorenie okna je možné pomocou vytvoreného detektora dobre detekovať, pričom však dochádza k veľkému počtu falošných detekcii otvorenia okna. Najlepšej upravenej úspešnosti a najmenšieho počtu false positives dosiahol detektor, vytvorený pomocou neurónovej siete (99,82 %).

	Prístup	**Úspešnosť [%]**	**True positive**	**False negative**	**True negative**	**False positive**
Hlboké učenie	klasický	91,70	323	2	196 519	17 815
	upravený	93,00	325	0	199 316	15 018
SVM	klasický	97,40	322	3	208 746	5 588
	upravený	98,63	325	0	211 403	2 931
Rozhodovací strom	klasický	95,82	325	0	205 367	8 967
	upravený	97,30	325	0	208 528	5 806
Náhodný les	klasický	96,26	325	0	206 297	8 037
	upravený	97,62	325	0	209 229	5 105
Bayesovská klasifikácia	klasický	98,44	325	0	194 724	3 099
	upravený	99,23	325	0	196 296	1 527
Neurónová sieť	klasický	98,76	322	3	211 679	2 655
	upravený	99,82	324	1	213 955	379

Tabuľka 7.13: Dosiahnuté úspešnosti klasifikátorov pri počiatočnom vyhodnotení v hlavnom testovacom prostredí (uvažovaná koncentrácia CO, vnútorná a vonkajšia teplota a vlhkosť)

Trénovacia fáza detektora opäť využívala iteratívny proces učenia, kedy bolo v každej iterácii nahradených 10 záznamov. Po vykonaní 60. iterácie bol dosiahnutý maximálny počet záznamov, ktoré je možné nahradiť v trénovacej sade. V prípade hlbokého učenia a Bayesovskej klasifikácie nebolo možné eliminovať všetky false positives. Hlavným dôvodom neúspešnej eliminácie bol veľký počet uvažovaných atribútov (1 773) a malý počet trénovacích dát. Najlepších výsledkov dosahoval detektor vytvorený pomocou náhodného lesa (100,00 %, viď tabuľka 7.14). Nastavením parametrov ostatných klasifikátorov by mohol byť znížený počet false negatives alebo false positives.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	99,20	323	2	212 619	1 715
	upravený	100,00	325	0	214 329	5
SVM	klasický	99,23	312	13	212 697	1 637
	upravený	100,00	316	9	214 334	0
Rozhodovací strom	klasický	99,17	312	13	212 572	1 762
	upravený	100,00	318	7	214 334	0
Náhodný les	klasický	99,32	320	5	212 872	1 462
	upravený	100,00	324	1	214 334	0
Bayesovská klasifikácia	klasický	99,41	320	5	210 643	1 252
	upravený	100,00	322	3	211 891	4
Neurónová sieť	klasický	99,16	320	5	212 532	1 802
	upravený	100,00	323	2	214 334	0

Tabuľka 7.14: Dosiahnuté úspešnosti klasifikátorov po 60. iterácii v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť)

Pre otestovanie natrénovaného detektora otvorenia okna boli využité dáta z hlavnej testovacej miestnosti (viď podsekcia 6.1.1), ktoré neboli použité v rámci iteratívneho trénovania. Testovacia sada sa skladala z 150 454 záznamov, definovaných na základe dát nazbieraných v období od začiatku februára 2019 do konca apríla 2019. Natrénovaný detektor pomocou náhodného lesa viď tabuľka 7.15) vykazoval najvyššiu úspešnosť (100,00 %). Výsledky získané všetkými klasifikátormi sú uvedené v tabuľke B.9, ktorá je dostupná v prílohe B. Neboli detekované len 3 otvorenia okna a zároveň boli detekované 4 falošné otvorenia okna.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Náhodný les	klasický	99,16	272	6	148 742	1 262
	upravený	100,00	275	3	150 000	4

Tabuľka 7.15: Úspešnosť náhodného lesa pri obecnom otestovaní v hlavnej testovacej miestnosti (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť)

Testovacia sada s 34 764 záznamami bola vytvorená pomocou dát z kancelárskej miestnosti, kde ich zber prebiehal počas apríla 2019. Najlepším klasifikátorom pre detekciu otvorenia okna bola neurónová sieť (99,98 %, viď tabuľka 7.16). Úspešnosti všetkých klasifikátorov sú zobrazené v tabuľke B.10, ktorá je dostupná v prílohe B. Použitý klasifikátor detekoval 6 false negatives a nedetekoval žiadne falošné otvorenie okna.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Neurónová sieť	klasický	99,97	0	8	34 740	4
	upravený	99,98	2	6	34 744	0

Tabuľka 7.16: Úspešnosť neurónovej siete pri obecnom otestovaní v kancelárskej miestnosti (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť)

Nasledujúcim testovacím prostredím bola obytná miestnosť 1, kde meranie prebiehalo od konca novembra 2018 do konca apríla 2019. Testovacia sada obsahovala 182 362 záznamov. Najvyššia úspešnosť bola dosiahnutá pomocou náhodného lesa (99,97 %, viď tabuľka 7.17). Úspešnosti všetkých klasifikátorov sú ukázané v tabuľke B.11, ktorá je dostupná v prílohe B. Počet false negatives bol 54 a počet false positives bol 9.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Náhodný les	klasický	99,58	204	115	181 139	656
	upravený	99,97	265	54	181 786	9

Tabuľka 7.17: Úspešnosť náhodného lesa pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť)

Testovacia sada, ktorá sa skladala z 39 524 záznamov, charakterizovala prostredie v obytnej miestnosti 2, kde bolo meranie vykonané počas apríla 2019. Detektor, ktorý dosahoval najvyššej úspešnosti, bol vytvorený pomocou neurónovej siete (99,93 %). Výsledky dosiahnuté všetkými klasifikátormi sú prezentované v tabuľke B.12, ktorá je dostupná v prílohe B. Z 22 otvorení okna boli detekované správne len 2 otvorenia okna. Počet nesprávne detekovaných otvorení okna bol 9.

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Neurónová sieť	klasický	99,91	0	22	39 467	13
	upravený	99,93	2	20	39 471	9

Tabuľka 7.18: Úspešnosť neurónovej siete pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť)

Hlavné príčiny chybne klasifikovaných udalostí sú väčšie objemy miestností a menšie plochy otvárateľných častí okien v porovnaní s hlavným testovacím prostredím. Ďalšou príčinou nízkej úspešnosti klasifikátorov môže byť umiestnenie senzora príliš ďaleko od okna. Podobne ako bolo uvedené u predchádzajúceho detektoru, väčší počet uvažovaných atribútov, neprispel k zlepšeniu úspešnosti.

Kapitola 8

Regulácia vzduchu v miestnosti

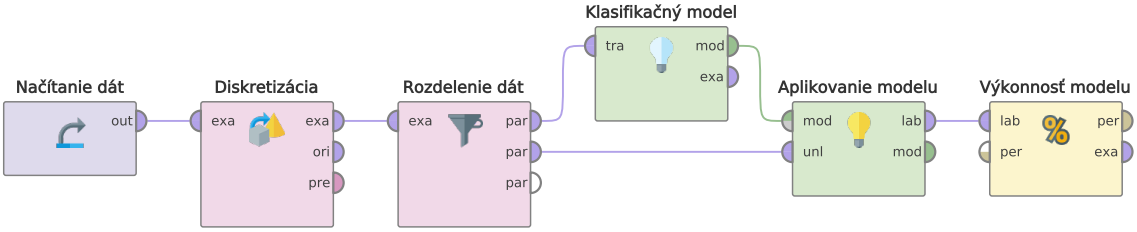
Jedným z cieľov práce bolo navrhnúť vhodný spôsob regulácie výmeny vzduchu v miestnosti pre zlepšenie kvality ovzdušia. Koncentrácia CO_2 je hlavným ukazovateľom kvality vnútorného ovzdušia, pričom vysoké hodnoty môžu spôsobiť zdravotné problémy. Zníženie koncentrácie CO_2 v miestnosti je možné dosiahnuť pomocou výmeny vzduchu, ktorá môže byť prirodzená alebo nútená. Pre reguláciu vzduchu v miestnosti bola využitá prirodzená výmena vzduchu, ktorá môže byť vykonaná otvorením okna.

Pri otvorení okna dôjde k postupnej výmene vzduchu v miestnosti, čo spôsobí pokles koncentrácie CO_2 na vonkajšiu úroveň, ktorá je približne od 350 ppm do 450 ppm. Na základe analýzy priebehu koncentrácie CO_2 počas 24 hodín bola vonkajšia úroveň stanovená na 434 ppm.

Pre odhad, za aký čas sa dostane koncentrácia CO_2 na určitú hodnotu, je možné použiť dva spôsoby. Prvým spôsobom je výpočet poklesu koncentrácie CO_2 podľa rovnice (2.9), ktorá pre výpočet vyžaduje hodnotu intenzity vetrania. Presný výpočet intenzity vetrania je problematický, a to hlavne z toho dôvodu, že je potrebné poznať objem miestnosti a intenzitu výmeny vzduchu. Intenzita výmeny vzduchu je veľmi závislá na vnútorných a vonkajších podmienkach počas otvorenia okna. Medzi parametre, ktoré ovplyvňujú pokles CO_2 najviac, patrí rozdiel medzi vnútornou a vonkajšou teplotou, sila vetra a umiestnenie budovy. Pre presné určenie intenzity výmeny vzduchu by bolo potrebné uvedené parametre presne odmerať, čo je problematické. Popísaný spôsob výpočtu je ideálny pre vzduchotechnické systémy, kde je prúdenie vzduchu konštantné.

Druhým spôsobom odhadu dĺžky vetrania je zistenie poklesu CO_2 pomocou exponenciálnej regresie, aplikovanej na namerané dáta. Výpočtom exponenciálnej regresie je možné určiť intenzitu vetrania a v prípade, že poznáme aj objem miestnosti je možné zistiť intenzitu výmeny vzduchu. Uvedeným postupom je možné spätne dopočítať intenzitu výmeny vzduchu, ku ktorej došlo počas otvorenia okna.

Pre odhad poklesu koncentrácie CO_2 bol vytvorený klasifikátor za pomoci programu RapidMiner. Vstupom klasifikátora je dátová sada, ktorá zahŕňa atribúty popísané v sekcii 6.3 (rozdiel vnútornej a vonkajšej teploty, vlhkosti, CO_2, exponenciálna regresia). Následne sa stĺpec, odpovedajúci hodnotám exponenciálnej regresie, diskretizuje do určitého počtu košov. Potom sa dátová sada rozdelí na dve časti, a to na trénovaciu (70 %) a testovaciu (30 %) sadu. Trénovacia sada ďalej slúži ako vstup klasifikačnej metódy. Na výsledný model sa aplikuje testovacia sada, na základe ktorej sa určí výkonnosť modelu. Zmienený postup je znázornený pomocou blokovej schémy, vytvorenej v programe RapidMiner na obrázku 8.1.



Obr. 8.1: Bloková schéma klasifikátora dĺžky vetrania v programe RapidMiner

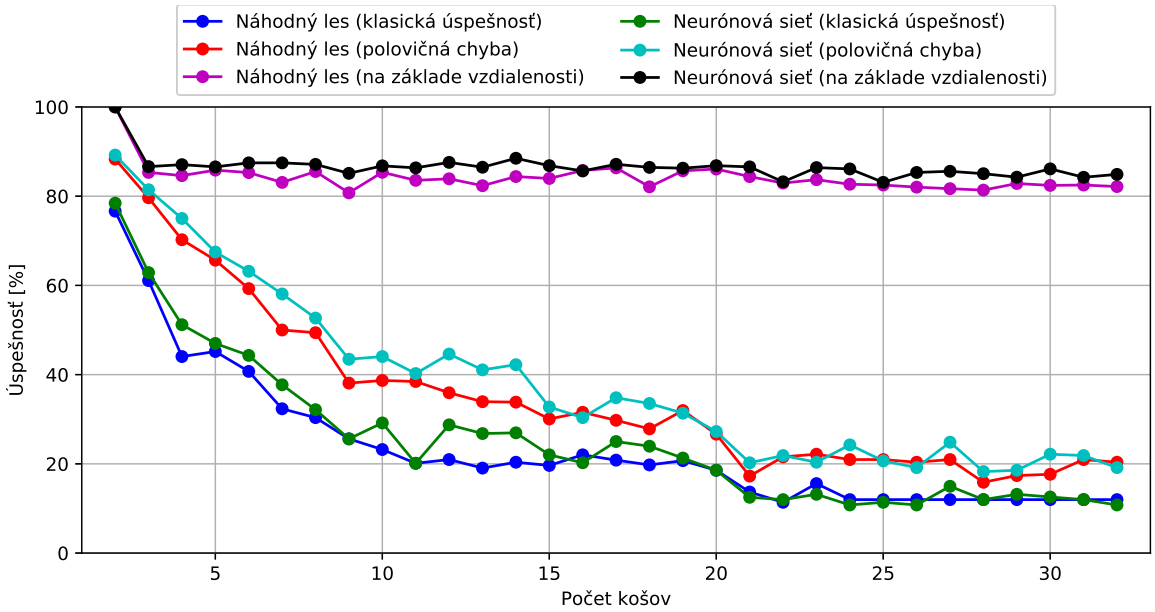
Meranie prebiehalo v hlavnej testovacej miestnosti od začiatku októbra 2018 do konca apríla 2019. Dátová sada sa skladala z 558 záznamov, pričom trénovaciu sadu tvorilo 392 záznamov a testovacia sada obsahovala 166 záznamov. Jednotlivé záznamy obsahovali triedu, do ktorej boli klasifikované. Triedy boli vytvorené na základe diskretizácie hodnôt intenzity výmeny vzduchu do košov. Klasifikácia prebiehala postupne do 2 až 32 košov, kedy sa testovala závislosť úspešnosti na počte uvažovaných košov (tried). Viacej ako 32 košov nebolo uvažované, pretože bolo meraním zistené, že sa už úspešnosť so zvyšujúcim sa počtom košov nemení. Model bol vytvorený za pomoci dvoch klasifikačných metód, a to náhodného lesa a neurónovej siete.

Dosiahnutá úspešnosť modelov bola vyhodnotená pomocou troch prístupov (viď obrázok 8.2), teda pomocou klasickej úspešnosti, polovičnej chyby a chyby na základe vzdialenosti. Najskôr bola uvažovaná klasická úspešnosť (viď rovnica (7.1)). Ďalšou možnosťou bola úspešnosť s polovičnou chybou, ktorá polovicu počtu záznamov, klasifikovaných do vedľajšieho koša, považovala ako správne klasifikované (true positives) záznamy. Dôvodom použitia polovičnej chyby bola eliminácia šumu v nameraných dátach, ktorá mohla spôsobiť miernu zmenu v poklese koncentrácie CO_2. Príčinou šumu v dátach môžu byť meniace sa podmienky (zmena sily vetra) a udalosti (príchod osoby do miestnosti) po otvorení okna. Vzhľadom k cieľovej doméne bolo zbytočné vykonávať príliš presnú klasifikáciu intenzity výmeny vzduchu, na základe ktorej je možné určiť požadovanú dĺžku vetrania. Odhad, ako dlho sa má vetrať, s presnosťou na sekundy pri reálnom nasadení neprináša užívateľom žiadnu pridanú hodnotu. Posledný spôsob vyhodnotenia úspešnosti využíval exponenciálnu regresiu nameraných dát po otvorení okna. Výpočet úspešnosti bral v úvahu vzdialenosť medzi priemernou hodnotou koša (intenzity výmeny vzduchu), kam daný záznam skutočne patril, a priemernou hodnotou koša, kam bol záznam, na základe vyššie uvedených atribútov, klasifikovaný. Vypočítaná vzdialenosť bola normalizovaná a ďalej považovaná za chybu pri klasifikácii jedného záznamu. Celková chyba bola potom definovaná ako súčet chyb pri klasifikácii všetkých záznamov. Výsledná úspešnosť (*uspesnost_podla_vzdialenosti*) bola definovaná pomocou vzťahu

$$uspesnost_podla_vzdialenosti = 1 - \frac{celkova_chyba}{pocet_zaznamov}. \tag{8.1}$$

Dosiahnuté výsledky sú prezentované pomocou obrázka 8.2. Úspešnosť uvažovaných klasifikátorov pri jednotlivých spôsoboch výpočtu úspešnosti je zrovnateľná. Dôvodom podobných výsledkov môže byť malý počet uvažovaných atribútov, alebo nedostatočný počet dát na trénovanie a testovanie. Klasická úspešnosť dosahovala najhorších výsledkov, pohybovala sa okolo 10 %–78 %, pričom sa približne od 20 košov, do ktorých boli záznamy klasifikované, už nemenila. Mierne lepších výsledkov, v rozmedzí 15 %–89 %, dosahovala polovičná úspešnosť, ktorá bola konštantná od zhruba rovnakého počtu košov. Najvyššia úspešnosť bola

dosiahnutá pomocou výpočtu na základe vzdialenosti, kedy nadobúdala hodnôt od 84 % do 100 %. Presné hodnoty úspešností klasifikátorov, zobrazených na obrázku 8.2, sú uvedené v tabuľke C.1, ktorá je dostupná v prílohe C.



Obr. 8.2: Závislosť úspešnosti klasifikátorov intenzity výmeny vzduchu na základe počtu košov a spôsobu výpočtu úspešnosti

Z výsledkov vyplýva, že voľba klasifikačnej metódy má zanedbateľný vplyv na úspešnosť klasifikácie intenzity výmeny vzduchu. Naproti tomu zvolený spôsob vyhodnotenia úspešnosti výrazne ovplyvňuje výslednú úspešnosť. Najlepším prístupom bolo vyhodnotenie, ktoré bralo do úvahy vzdialenosti medzi skutočnými a odhadnutými hodnotami intenzity výmeny vzduchu. Vzdialenosť vhodne reflektuje chybu klasifikácie do vedľajších tried. S rastúcou chybou klasifikácie, teda s rastúcou vzdialenosťou medzi skutočnou a odhadnutou triedou, sa úspešnosť znižuje. Spomenutý prístup využíval fyzikálneho modelu poklesu koncentrácie CO_2 počas vetrania, ktorý dokáže na základe aktuálne nameraných dát predpovedať hodnotu intenzity výmeny vzduchu.

Kapitola 9

Záver

Cieľom práce bola analýza kvality ovzdušia v kancelárskych a obytných priestoroch. Všetky ciele práce boli splnené. Najskôr bola naštudovaná problematika kvality ovzdušia vo vnútorných priestoroch, ktorá je predovšetkým ovplyvnená koncentráciou CO_2. Medzi hlavnými faktory ovplyvňujúce koncentráciu CO_2 patria ľudia a množstvo vymieňaného vzduchu. Ďalej boli popísané vlastnosti CO_2 a jeho účinky pri vyšších koncentráciách, ktoré môžu spôsobovať únavu, bolesti hlavy a hrdla. Okrem problematiky kvality ovzdušia bol naštudovaný proces získavania znalostí z databáz, so zameraním sa na najpoužívanejšie klasifikačné a predikčné metódy, a tiež boli popísané spôsoby redukcie dimenzionality. Následne boli naštudované princípy merania koncentrácie CO_2, architektúra BeeeOn systému a boli zvolené vhodné senzory pre dlhodobé meranie.

Cieľom dlhodobého merania bolo získať dostatočné množstvo dát pre vytvorenie klasifikátorov na detekciu otvorenia okna a určenie intenzity výmeny vzduchu. Meranie prebiehalo v štyroch testovacích miestnostiach, z nich tri boli obytné a jedna kancelárska, medzi merané veličiny patrila koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť.

Pre detekciu otvorenia okna boli vytvorené tri rôzne detektory, ktoré uvažovali rôzne sady atribútov a boli trénované na dátach získaných z hlavnej testovacej miestnosti. Klasifikácia dát bola vykonaná pomocou hlbokého učenia, SVM, rozhodovacieho stromu, náhodného lesa, Bayesovskej klasifikácie a neurónovej siete. Prvý detektor detekoval otvorenie okna len na základe koncentrácie CO_2. Trénovacia sada sa skladala z 1 140 záznamov, ktoré boli reprezentované dátami, meranými od začiatku októbra 2018 do začiatku februára 2019. Testovacia sada obsahovala 237 974 záznamov, ktoré boli charakterizované dátami nameranými za obdobie od začiatku októbra 2018 do začiatku februára 2019. Každý záznam obsahoval 443 atribútov. Pri počiatočnom vyhodnotení klasifikátorov dosiahol najlepších výsledkov SVM klasifikátor, a to 98,91 %. Po vykonaní 58 iterácii dosiahlo najvyššej úspešnosti hlboké učenie, a to 100 %, pričom boli eliminované všetky false negatives a false positives. Detekcia otvorenia okna na základe ďalších veličín, a to vnútornej a vonkajšej teploty a vlhkosti, nedosahovala lepších výsledkov, nepodarilo sa eliminovať všetky false positives a false negatives.

Čo sa týka vyhodnotenia obecnosti, najvyššej úspešnosti bolo dosiahnuté za pomoci náhodného lesa (100,00 %) alebo neurónovej siete (100,00 %). Bolo zistené, že hlboké učenie je veľmi ovplyvnené charakterom dát, ktoré sa použijú pre trénovanie modelu. V prípade použitia modelu, vytvoreného na základe hlbokého učenia, v inom než trénovacom prostredí, dochádzalo v porovnaní s ostatnými modelmi, vytvorenými pomocou ostatných klasifikačných metód, k detekcii výrazne väčšieho počtu false positives.

Na základe dosiahnutých výsledkov bolo zistené, že nie je možné vytvoriť detektor otvorenia okna na základe dát nameraných v jednej miestnosti a použiť ho na detekciu otvorenia okna v inej miestnosti. Detektor musí byť naučený na parametre danej miestnosti.

Pre reguláciu vzduchu v miestnosti bolo zvolené prirodzené vetranie. Hlavným parametrom, ktorý ovplyvňuje výmenu vzduchu v miestnosti, je intenzita výmenu vzduchu, ktorá bola využitá pre klasifikáciu dát. Dátová sada zahŕňala 558 záznamov, vytvorených z dát získaných za obdobie od začiatku októbra 2018 do polovice mája 2019. Trénovacia sada obsahovala 70 % záznamov z dátovej sady a testovaciu sadu tvorilo 30 % záznamov. Bola uvažovaná neurónová sieť a náhodný les, pričom bola skúmaná závislosť úspešnosti klasifikátorov intenzity výmeny vzduchu na základe počtu košov (intervalov intenzít výmeny vzduchu) a prístupu k výpočtu úspešnosti.

Vyhodnotenie úspešnosti klasifikátorov bolo vykonané tromi rôznymi spôsobmi. Výsledky ukázali, že sa úspešnosť od určitého počtu košov nemení. Najlepších výsledkov, úspešnosti od 84 % do 100 %, dosahoval prístup, ktorý bral v úvahu vzdialenosť medzi priemernou hodnotou koša (intenzity výmeny vzduchu), kam daný záznam skutočne patril, a priemernou hodnotou koša, kam bol záznam klasifikovaný. Dôvodom vysokej úspešnosti vyhodnotenia na základe vzdialenosti bolo využitie modelu, ktorý bol založený na fyzikálnej podstate výmeny vzduchu. Pomocou modelu bolo možné predpovedať hodnotu intenzity výmeny vzduchu, ktorá slúži pre jednoduchý výpočet doby, za ktorú koncentrácia CO_2 dosiahne zvolenú hodnotu.

V práci je možné pokračovať zbieraním ďalších dát spojených s poklesom alebo nárastom CO_2, ktoré je možné následne použiť pre vytvorenie presnejších klasifikátorov a ich otestovanie. S narastajúcim počtom nameraných dát bude možné vytvoriť kvalitnejší klasifikátor intenzity výmeny vzduchu. Ďalej je možné navrhnúť systém, ktorý bude s rastúcim množstvom nazbieraných dát spresňovať model. Každá miestnosť by mohla využívať model, vytvorený pomocou zautomatizovania navrhnutého systému, ktorý sa skladá z anotácie dát a postupného spresňovania modelu na základe nameraných dát v danej miestnosti.

Literatúra

[1] AGROBAC: Plícní objemy a jejich měření. [Online; navštívené 28.04.2019]. URL https://galenus.cz/clanky/biochemie/biochemie-fyziologie-plicni-objemy

[2] BBC: What happens during photosynthesis? [Online; navštívené 25.04.2019]. URL http://www.bbc.co.uk/schools/gcsebitesize/science/add_ocr_21c/life_processes/plantfoodrev1.shtml

[3] Bouckaert, R. R.; Frank, E.; Hall, M.; aj.: WEKA Manual for Version 3-7-13. November 2015, [Online; navštívené 20.04.2019]. URL https://sourceforge.net/projects/weka/postdownload?source=dlp

[4] Braniš, M.: *Atmosféra a klima. Aktuální otázky znečištění ovzduší.* Karolinum, 2009, ISBN 9788024615981, 351 s.

[5] Burges, C. J.: A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, ročník 2, č. 2, Jun 1998: s. 121–167, ISSN 1573-756X.

[6] Bušta, K.; Wagner, Z.: Oxid uhličitý a možnosti jeho využití I – fyzikální a chemické vlastnosti a jejich využití v praxi. *TZB-info*, Apríl 2012, ISSN 1801-4399, [Online; navštívené 28.02.2019]. URL https://oze.tzb-info.cz/teorie-obnovitelna-energie/8492-oxid-uhlicity-a-moznosti-jeho-vyuziti-i

[7] Centre, L. H.: *Sick Building Syndrome: Causes Effects and Control.* London Hazards Centre True Limited, Jún 1990, ISBN 0948974060, 94 s.

[8] to Chemical Hazards Carbon dioxide, C. N. P. G.: CDC - NIOSH Pocket Guide to Chemical Hazards - Carbon dioxide. 2016, [Online; navštívené 20.03.2019]. URL https://www.cdc.gov/niosh/npg/npgd0103.html

[9] Cifrinec, I.: Větrání bytových domů - Základy teorie větrání. *TZB-info*, Máj 2010, ISSN 1801-4399, [Online; navštívené 19.02.2019]. URL https://vetrani.tzb-info.cz/vetrani-bytovych-domu/6507-vetrani-bytovych-domu-zaklady-teorie-vetrani

[10] Doležílková, H.: Kvalita vnějšího a vnitřního vzduchu. *TZB-info*, Máj 2010, ISSN 1801-4399, [Online; navštívené 28.02.2019]. URL https://vetrani.tzb-info.cz/vnitrni-prostredi/6486-kvalita-vnejsiho-a-vnitrniho-vzduchu

[11] Drucker, H.: Improving Regressors Using Boosting Techniques. In *ICML*, 08 1997.

[12] Hazucha, J.; Bárta, J.: *Nucené větrání s možností rekuperace odpadního tepla v objektech pro vzdělávání.* Státní fond životního prostředí ČR, Január 2010, 44 s.

[13] Home, I.: CO2 Temperature Relative Humidity Sensor. [Online; navštívené 20.03.2019].
URL https://www.iqhome.org/sensors/SN-THC-02

[14] Hrabal, P.: Učebnice chemie - oxid uhličitý. [Online; navštívené 20.03.2019].
URL http://ucebnicechemie.wz.cz/index.php?sloucenina=oxid_uhlicity

[15] Jones, A.: Indoor air quality and health. *Atmospheric Environment*, ročník 33, č. 28, 1999: s. 4535–4564, ISSN 13522310, [Online; navštívené 20.03.2019].
URL http://www.sciencedirect.com/science/article/pii/S1352231099002721

[16] Kaushik, S.: Introduction to Feature Selection methods with an example. December 2016, [Online; navštívené 28.04.2019].
URL https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/

[17] Kotlík, B.: Vnitřní prostředí (Urbanismus a bydlení) - hygienické aspekty. [Online; navštívené 12.03.2019].
URL http://www.szu.cz/tema/zivotni-prostredi/urbanismus-a-bydleni-hygienicke-aspekty

[18] Kratěnová, J.; Puklová, V.: Venkovní ovzduší. Január 2015, [Online; navštívené 09.02.2019].
URL http://www.szu.cz/tema/zivotni-prostredi/venkovni-ovzdusi-2

[19] Laußmann, D.; Helm, D.: Air Change Measurements Using Tracer Gases: Methods and Results. Significance of air change for indoor air quality. In *Chemistry, Emission Control, Radioactive Pollution and Indoor Air Quality, Chapter 14*, Robert Koch-Institut, Epidemiologie und Gesundheitsberichterstattung, 2011, str. 42, [Online; navštívené 19.02.2019].
URL https://edoc.rki.de/bitstream/handle/176904/973/27OA2TGPWvv0.pdf

[20] Li, J.; Cheng, K.; Wang, S.; aj.: Feature Selection: A Data Perspective. *ACM Computing Surveys*, ročník 50, 01 2016.

[21] Mathauserová, Z.: Význam větrání v budovách. [Online; navštívené 22.03.2019].
URL http://www.cklop.cz/2014/02_Mathauserova_Vyznam_vetrani.pdf

[22] MEFANET, síť lékařských fakult ČR a SR: Hemoglobin a jeho deriváty. Január 2018, [Online; navštívené 20.03.2019].
URL https://www.wikiskripta.eu/w/Hemoglobin_a_jeho_deriváty_(LF_MU)

[23] Pal, M.: Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, ročník 26, č. 1, 2005: s. 217–222.

[24] Polikar, R.: *Pattern Recognition.* American Cancer Society, 2006, ISBN 9780471740360.

[25] Protronix: Kombinované čidlo CO2, RH, T s IQRF. [Online; navštívené 15.02.2019].
URL https://www.careforair.eu/wp-content/uploads/2017/12/um-NLII-CO2RHT-IQRF-cz-V3-171211.pdf

[26] Protronix: Pracujete ve zdravém prostředí? - Měření oxidu uhličitého v budovách. *TZB-info*, August 2009, ISSN 1801-4399, [Online; navštívené 15.02.2019].
URL https://vetrani.tzb-info.cz/vnitrni-prostredi/5827-pracujete-ve-zdravem-prostredi

[27] Rangra, K.; Bansal, K.: Comparative Study of Data Mining Tools. *International journal of advanced research in computer science and software engineering*, ročník 4, č. 6, 2014.

[28] Česká Republika: Vyhláška, kterou se mění vyhláška č. 268/2009 Sb., o technických požadavcích na stavby. In *Sbírka zákonů*, Tlačiareň Ministerstva vnútra, Február 2012, str. 6, [Online; navštívené 23.12.2018].
URL https://aplikace.mvcr.cz/sbirka-zakonu/ViewFile.aspx?type=c&id=6102

[29] Rojas, R.: *Neural Networks: A Systematic Introduction*. Springer-Verlag, 01 1996, 453 s.

[30] Slotová, K.: Faktory vnútorného ovzdušia budov a ich vplyv na zdravie obyvateľov . 2010, [Online; navštívené 15.03.2019].
URL http://www.vzbb.sk/sk/urad/narodne_centra/nrc_vo/faktory_vnut_ovzdusia101001.pdf

[31] TZB-info: Složení atmosferického vzduchu. [Online; navštívené 20.03.2019].
URL https://www.tzb-info.cz/tabulky-a-vypocty/74-slozeni-atmosferickeho-vzduchu

[32] University of Ljubljana: Data Mining Fruitful and Fun. Online, [Online; navštívené 19.02.2019].
URL https://orange.biolab.si/

[33] Vafek, Z.; Safety, D.: Možnosti měření oxidu uhličitého – měřicí přístroje a čidla. *TZB-info*, Marec 2016, ISSN 1801-4399, [Online; navštívené 19.02.2019].
URL https://vetrani.tzb-info.cz/vnitrni-prostredi/13910-moznosti-mereni-oxidu-uhliciteho-merici-pristroje-a-cidla

[34] Vidim, J.: Měření kvality vzduchu a CO2 v interiérech. August 2013, [Online; navštívené 19.02.2019].
URL https://www.asb-portal.cz/tzb/vetrani-a-klimatizace/mereni-kvality-vzduchu-a-co2-v-interierech

[35] Zendulka, J.; Kunc, M.; Stryka, L.: Získávání znalostí z databází ZZN. September 2009, [Online; navštívené 19.02.2019].

[36] Zikán, Z.: Oxid uhličitý - utajený nepřítel. *TZB-info*, Február 2011, ISSN 1801-4399, [Online; navštívené 19.02.2019].
URL https://vetrani.tzb-info.cz/vnitrni-prostredi/7132-oxid-uhlicity-utajeny-nepritel

[37] Zmrhal, V.: Požadavky na větrání obytných budov dle ČSN EN 15 665/Z1. *TZB-info*, Január 2012, ISSN 1801-4399, [Online; navštívené 19.02.2019]. URL https://vetrani.tzb-info.cz/normy-a-pravni-predpisy-vetrani-klimatizace/8239-pozadavky-na-vetrani-obytnych-budov-dle-csn-en-15-665-z1

[38] Úřad pro technickou normalizaci, metrologii a státní zkušebnictví: *ČSN EN 16798-3 – Energetická náročnost budov – Větrání budov – Část 3: Větrání nebytových budov – Základní požadavky na větrací a klimatizační zařízení*. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2018.

Príloha A

Obsah priloženého pamäťového média

Priložené pamäťové médium obsahuje:

- `processes` – procesy vytvorené v programe RapidMiner pre učenie a testovanie detektora a prediktora,
- `src` – všetky súbory, ktoré sú potrebné pre predspracovanie nameraných dát a vytvorenie požadovaných CSV súborov,
- `src_Tisovcik` – všetky zdrojové súbory, ktoré vytvoril autor tejto diplomovej práce,
- `dp_latex` – zdrojové kódy textu diplomovej práce,
- `xtisov00.pdf` – text diplomovej práce,
- `README.md` – inštrukcie pre spustenie zdrojových kódov.

Príloha B

Úspešnosti detektorov pri vyhodnotení obecnosti

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	99,17	348	3	187 508	1 568
	upravený	99,99	349	2	189 058	18
SVM	klasický	99,02	323	28	187 255	1 821
	upravený	99,98	332	19	189 055	21
Rozhodovací strom	klasický	99,39	320	31	187 949	1 127
	upravený	99,99	332	19	189 072	4
Náhodný les	klasický	99,13	338	13	187 434	1 642
	upravený	99,99	343	8	189 071	5
Bayesovská klasifikácia	klasický	99,32	337	14	185 898	1 252
	upravený	99,99	344	7	187 137	13
Neurónová sieť	klasický	99,05	343	8	187 291	1 785
	upravený	99,99	347	4	189 060	16

Tabuľka B.1: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v hlavnom testovacom prostredí (uvažovaná len koncentrácia CO_2)

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	93,87	1	9	39 240	2 553
	upravený	93,93	8	2	39 256	2 537
SVM	klasický	99,98	0	10	41 793	0
	upravený	99,98	0	10	41 793	0
Rozhodovací strom	klasický	99,97	0	10	41 792	1
	upravený	99,98	1	9	41 793	0
Náhodný les	klasický	99,98	0	10	41 793	0
	upravený	99,98	0	10	41 793	0
Bayesovská klasifikácia	klasický	99,98	0	10	41 766	0
	upravený	99,98	0	10	41 766	0
Neurónová sieť	klasický	99,93	0	10	41 773	20
	upravený	99,95	3	7	41 778	15

Tabuľka B.2: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v kancelárskej miestnosti (uvažovaná len koncentrácia CO_2)

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	97,71	315	126	276 570	6 371
	upravený	98,27	428	13	278 045	4 896
SVM	klasický	99,44	163	278	281 643	1 298
	upravený	99,94	325	116	282 889	52
Rozhodovací strom	klasický	99,70	131	310	282 410	531
	upravený	99,95	297	144	282 931	10
Náhodný les	klasický	99,68	151	290	282 333	608
	upravený	99,94	164	177	282 939	2
Bayesovská klasifikácia	klasický	99,84	5	433	282 723	20
	upravený	99,85	11	427	282 740	3
Neurónová sieť	klasický	99,36	258	183	281 300	1 641
	upravený	99,96	406	35	282 871	70

Tabuľka B.3: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná len koncentrácia CO_2)

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	94,90	4	26	56 556	3 014
	upravený	94,98	16	14	56 591	2 979
SVM	klasický	99,92	0	30	59 553	17
	upravený	99,93	2	28	59 557	13
Rozhodovací strom	klasický	99,93	0	30	59 558	12
	upravený	99,96	7	23	59 569	1
Náhodný les	klasický	99,94	0	30	59 565	5
	upravený	99,95	1	29	59 568	2
Bayesovská klasifikácia	klasický	99,95	0	30	59 403	1
	upravený	99,95	0	30	59 403	1
Neurónová sieť	klasický	99,78	2	28	59 467	103
	upravený	99,84	11	19	59 492	78

Tabuľka B.4: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná len koncentrácia CO_2)

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	99,00	348	3	187 194	1 882
	upravený	99,99	351	0	189 049	27
SVM	klasický	99,10	340	11	187 388	1 688
	upravený	99,99	344	7	189 070	6
Rozhodovací strom	klasický	99,13	320	31	187 468	1 608
	upravený	99,99	331	20	189 070	6
Náhodný les	klasický	99,18	342	9	187 539	1 537
	upravený	100,00	347	4	189 072	4
Bayesovská klasifikácia	klasický	99,31	344	7	187 165	1 303
	upravený	99,99	345	6	188 457	11
Neurónová sieť	klasický	98,98	346	5	187 150	1 926
	upravený	99,99	349	2	189 063	13

Tabuľka B.5: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť)

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	99,51	0	10	41 598	195
	upravený	99,54	4	6	41 606	187
SVM	klasický	99,97	0	10	41 792	1
	upravený	99,98	1	9	41 793	0
Rozhodovací strom	klasický	99,97	0	10	41 789	4
	upravený	99,98	1	9	41 793	0
Náhodný les	klasický	99,97	0	10	41 790	3
	upravený	99,98	1	9	41 793	0
Bayesovská klasifikácia	klasický	99,98	0	10	41 777	0
	upravený	99,98	0	10	41 777	0
Neurónová sieť	klasický	99,97	0	10	41 790	3
	upravený	99,98	1	9	41 793	0

Tabuľka B.6: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v kancelárskej miestnosti (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť)

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	98,24	281	74	212 287	3 740
	upravený	98,95	349	6	213 770	2 257
SVM	klasický	99,32	259	96	214 648	1 379
	upravený	99,89	315	40	215 821	206
Rozhodovací strom	klasický	99,12	279	76	214 201	1 826
	upravený	99,90	324	31	215 831	196
Náhodný les	klasický	99,57	248	107	215 193	834
	upravený	99,96	306	49	216 000	27
Bayesovská klasifikácia	klasický	99,83	4	351	215 982	12
	upravený	99,84	6	349	215 992	2
Neurónová sieť	klasický	99,08	284	71	214 113	1 914
	upravený	99,81	337	18	215 627	400

Tabuľka B.7: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť)

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	99,19	0	30	59 120	450
	upravený	99,24	12	18	59 138	432
SVM	klasický	99,95	0	30	59 570	0
	upravený	99,95	0	30	59 570	0
Rozhodovací strom	klasický	99,92	0	30	59 551	19
	upravený	99,92	0	30	59 551	19
Náhodný les	klasický	99,95	0	30	59 570	0
	upravený	99,95	0	30	59 570	0
Bayesovská klasifikácia	klasický	99,95	0	30	59 506	1
	upravený	99,95	0	30	59 506	1
Neurónová sieť	klasický	99,94	0	30	59 564	6
	upravený	99,95	2	28	59 566	4

Tabuľka B.8: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná koncentrácia CO_2, vnútorná teplota a vlhkosť)

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	98,90	277	1	148 345	1 659
	upravený	99,88	278	0	149 823	181
SVM	klasický	99,08	268	10	148 628	1 376
	upravený	99,99	271	7	149 992	12
Rozhodovací strom	klasický	98,97	271	7	148 467	1 537
	upravený	99,99	276	2	149 994	10
Náhodný les	klasický	99,16	272	6	148 742	1 262
	upravený	100,00	275	3	150 000	4
Bayesovská klasifikácia	klasický	99,29	272	6	147 448	1 051
	upravený	99,99	273	5	148 488	11
Neurónová sieť	klasický	98,98	275	3	148 470	1 534
	upravený	99,99	276	2	149 989	15

Tabuľka B.9: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v hlavnom testovacom prostredí (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť)

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	99,66	0	8	34 635	109
	upravený	99,69	3	5	34 642	102
SVM	klasický	99,97	0	8	34 743	1
	upravený	99,98	1	7	34 744	0
Rozhodovací strom	klasický	99,95	0	8	34 733	11
	upravený	99,98	4	4	34 741	3
Náhodný les	klasický	99,97	0	8	34 742	2
	upravený	99,98	1	7	34 744	0
Bayesovská klasifikácia	klasický	99,98	0	8	34 716	0
	upravený	99,98	0	8	34 716	0
Neurónová sieť	klasický	99,97	0	8	34 740	4
	upravený	99,98	2	6	34 744	0

Tabuľka B.10: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v kancelárskej miestnosti (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť)

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	97,54	267	52	177 362	4 433
	upravený	98,38	313	6	178 846	2 949
SVM	klasický	99,29	219	100	180 601	1 194
	upravený	99,96	286	33	181 752	43
Rozhodovací strom	klasický	99,04	233	86	180 135	1 660
	upravený	99,82	299	20	181 492	303
Náhodný les	klasický	99,58	204	115	181 139	656
	upravený	99,97	265	54	181 786	9
Bayesovská klasifikácia	klasický	99,82	15	303	180 223	14
	upravený	99,83	19	299	180 235	2
Neurónová sieť	klasický	99,11	246	73	180 255	1540
	upravený	99,87	300	19	181 586	209

Tabuľka B.11: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 1 (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť)

	Prístup	Úspešnosť [%]	True positive	False negative	True negative	False positive
Hlboké učenie	klasický	99,08	0	22	39 140	340
	upravený	99,11	4	18	39 146	334
SVM	klasický	99,92	0	22	39 471	9
	upravený	99,92	0	22	39 471	9
Rozhodovací strom	klasický	99,91	0	22	39 466	14
	upravený	99,91	1	21	39 467	13
Náhodný les	klasický	99,94	0	22	39 480	0
	upravený	99,94	0	22	39 480	0
Bayesovská klasifikácia	klasický	99,93	0	22	39 272	6
	upravený	99,93	0	22	39 272	6
Neurónová sieť	klasický	99,91	0	22	39 467	13
	upravený	99,93	2	20	39 471	9

Tabuľka B.12: Úspešnosť všetkých klasifikátorov pri obecnom otestovaní v obytnej miestnosti 2 (uvažovaná koncentrácia CO_2, vnútorná a vonkajšia teplota a vlhkosť)

Príloha C

Úspešnosť klasifikátorov pri určení intenzity výmeny vzduchu

	Náhodný les			**Neurónová sieť**		
Počet košov	klasická úspešnosť [%]	polovičná úspešnosť [%]	úspešnosť na základe vzdialenosti [%]	klasická úspešnosť [%]	polovičná úspešnosť [%]	úspešnosť na základe vzdialenosti [%]
2	76,65	88,32	100,0	78,44	89,22	100,00
3	61,08	79,64	85,33	62,87	81,44	86,63
4	44,05	70,24	84,61	51,19	75,00	87,07
5	45,18	65,66	85,84	46,99	67,47	86,57
6	40,72	59,28	85,29	44,31	63,17	87,47
7	32,34	50,00	83,09	37,72	58,08	87,47
8	30,36	49,40	85,53	32,14	52,68	87,13
9	25,60	38,10	80,75	25,60	43,45	85,12
10	23,21	38,69	85,36	29,17	44,05	86,80
11	20,12	38,46	83,55	20,12	40,24	86,35
12	20,96	35,93	83,90	28,74	44,61	87,58
13	19,05	33,93	82,35	26,79	41,07	86,49
14	20,36	33,83	84,40	26,95	42,22	88,50
15	19,64	30,06	83,97	22,02	32,74	86,84
16	22,02	31,55	85,77	20,24	30,36	85,66
17	20,83	29,76	86,35	25,00	34,82	87,15
18	19,76	27,84	82,10	23,95	33,53	86,46
19	20,71	31,95	85,69	21,30	31,36	86,28
20	18,56	26,65	86,09	18,56	27,25	86,84
21	13,68	17,26	84,39	12,50	20,24	86,57
22	11,38	21,56	82,94	11,98	21,86	83,22
23	15,57	22,16	83,70	13,17	20,36	86,41
24	11,98	20,96	82,68	10,78	24,25	86,11
25	11,98	20,96	82,50	11,38	20,66	83,10
26	11,98	20,36	82,03	10,78	19,16	85,32
27	11,98	20,96	81,68	14,97	24,85	85,58

28	11,98	15,87	81,36	11,98	18,26	85,04
29	11,98	17,37	82,84	13,17	18,56	84,26
30	11,98	17,66	82,42	12,57	22,16	86,12
31	11,98	20,96	82,50	11,98	21,86	84,24
32	11,98	20,36	82,16	10,78	19,16	84,89

Tabuľka C.1: Úspešnosť klasifikátorov pri určení intenzity výmeny vzduchu v závislosti na počte košov a spôsobu vyhodnotenia úspešnosti