

**Jihočeská univerzita v Českých Budějovicích  
Přírodovědecká fakulta**

## **Analýza dat vytápěných budov**

Bakalářská práce

**Martin Zeman**

Vedoucí práce: Ing. Miroslav Skrbek, Ph.D.

České Budějovice 2019

## **Bibliografické údaje**

Zeman, M., 2019: Analýza dat vytápěných budov. [Data analysis of heated buildings. Bc. Thesis, in Czech.] – 63 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

## **Anotace**

Tato bakalářská práce se zabývá analýzou dat naměřených v topných soustavách inteligentních budov, jejichž cílem je rešerše vhodných metod pro modelování chování topné soustavy a detekci anomálií. Pro analýzu byla poskytnuta data společností zabývajících se automatizací vytápěných inteligentních budov. V analýze řešení se věnuji charakteristice dat a zónovému systému, z něhož data vznikla. V návrhu řešení popisují export dat z databáze, selekci dat, předzpracování dat, transformaci dat, dolování znalostí z dat a samotnou vizualizaci dat. V rámci bakalářská práce bylo vytvořeno několik programů, jejichž implementace je popsána v implementaci řešení. Poslední část práce je zaměřena na experimenty s daty. K samotným experimentům byly použity zejména neuronové sítě, korelace, spojnicové grafy a teplotní mapy.

## **Klíčová slova**

Analýza dat, dolování dat, inteligentní budovy, Jupyter Notebook, korelace, neuronové sítě, Python, RapidMiner

## **Annotation**

This bachelor thesis deal with analysis of data measured in heating systems of intelligent buildings. The target of this bachelor thesis is to search for suitable methods for modeling the behavior of the heating system and to detect anomalies. The solution design describes data export from database, data selection, data pre-processing, data transformation, data mining and data visualization. In the bachelor thesis were created several programs, whose implementation is described in the implementation of the solution. The last part is focused on experiments with data. Especially neural networks, correlations, line graphs and temperature maps were used for the experiments.

## **Keywords**

Correlation, data analysis, data mining, intelligent building, Jupyter Notebook, neural network, Python, RapidMiner

## **Prohlášení**

Prohlašuji, že svojí bakalářskou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury. Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě, elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejich internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, dne 17. 04. 2019

---

Martin Zeman

## **Poděkování**

Tímto bych rád poděkoval panu Ing. Miroslavovi Skrbkovi, Ph.D. za odborné vedení práce, věcné připomínky, cenné rady a vstřícnost při konzultacích a vypracování bakalářské práce.

# Obsah

<b>1.</b>	<b>ÚVOD A CÍLE PRÁCE .....</b>	<b>9</b>
1.1	Úvod .....	9
1.2	Cíle práce .....	10
<b>2.</b>	<b>ANALÝZA ŘEŠENÍ.....</b>	<b>11</b>
2.1	Základní pojmy .....	11
2.2	Data .....	12
2.3	Vznik dat.....	13
2.3.1	Zónový systém.....	13
2.3.2	Části zónového systému .....	13
2.3.3	Princip zónového systému .....	14
2.4	Struktura dat.....	15
2.4.1	Databáze .....	15
2.4.2	Data z individuální regulace .....	15
2.4.3	Data z kotelny .....	17
<b>3.</b>	<b>NÁVRH ŘEŠENÍ.....</b>	<b>18</b>
3.1	Export dat z databáze.....	18
3.1.1	Obecný popis .....	18
3.1.2	Proces exportu dat .....	19
3.2	Aplikace pro zpracování dat .....	21
3.2.1	Anaconda .....	21
3.2.2	Jupyter Notebook.....	21
3.2.3	RapidMiner.....	22
3.3	Selekce dat .....	22
3.3.1	Obecný popis .....	22
3.3.2	Načtení dat.....	22
3.3.3	Rozdělení dat .....	23
3.3.4	Uložení dat.....	24
3.4	Předzpracování dat.....	24
3.4.1	Obecný popis .....	24

3.4.2	Načtení a uložení dat .....	24
3.4.3	Formát času.....	25
3.4.4	Funkce pro analýzu souboru.....	25
3.4.5	Funkce pro všechny dny .....	28
3.4.6	Závěr předzpracování dat .....	29
<b>3.5</b>	<b>Transformace dat.....</b>	<b>29</b>
3.5.1	Obecný popis .....	29
3.5.2	Doplnění atributů.....	29
3.5.3	Selekce hodnot atributů .....	30
3.5.4	Selekce atributů .....	31
<b>3.6</b>	<b>Dolování znalostí .....</b>	<b>32</b>
3.6.1	Obecný popis .....	32
3.6.2	Korelace.....	32
3.6.3	Umělé neuronové sítě .....	34
<b>3.7</b>	<b>Vizualizace .....</b>	<b>36</b>
3.7.1	Obecný popis .....	36
3.7.2	Teplotní mapa.....	36
3.7.3	Spojnicový graf.....	37
<b>4.</b>	<b>IMPLEMENTACE ŘEŠENÍ.....</b>	<b>37</b>
<b>5.</b>	<b>EXPERIMENTÁLNÍ VÝSLEDKY .....</b>	<b>38</b>
5.1	Přetápění a promrzání zón.....	38
5.2	Korelace jednotlivých zón .....	42
5.3	Chování zóny .....	45
5.4	Predikce příznaku zóny .....	49
<b>6.</b>	<b>ZÁVĚR.....</b>	<b>55</b>
<b>7.</b>	<b>LITERÁRNÍ PŘEHLED.....</b>	<b>57</b>
7.1	Knižní publikace .....	57
7.2	Manuály .....	57
7.3	Webové publikace .....	57
<b>8.</b>	<b>SEZNAM OBRÁZKŮ .....</b>	<b>60</b>

<b>9.</b>	<b>SEZNAM TABULEK.....</b>	<b>62</b>
<b>10.</b>	<b>SEZNAM PŘÍLOH.....</b>	<b>63</b>



# 1. Úvod a cíle práce

## 1.1 Úvod

Dnešní moderní a dostupné technologie, které označujeme jako „smart“, v překladu inteligentní či chytrý, nám v mnoha ohledech usnadňují život a stávají se nedílnou součástí našich životů. Dnes toto označení nenesou jen chytré zařízení, ale můžeme se setkávat i se chytrými budovami či dokonce chytrými městy, která nám poskytují tyto technologie. Využívání moderních technologií může být prostředkem k lepšímu životu. Popularita chytrých technologií navíc stále roste, ať už z důvodu stále se zvyšujících cen energií, ohleduplnosti k životnímu prostředí nebo, již zmiňované dostupnosti a vývoji technologií.

Dnešní chytré technologie, které z budov vytvářejí inteligentní budovy nám poskytují automatizaci technologií, jako je řízení vytápění, větrání, chlazení, osvětlení a mnoho dalších. Z důvodu zlepšení těchto systémů a dosažení vyšší úspory energie vzniká potřeba analýzy výstupních dat těchto systémů.

V této bakalářské práci se zabývám analýzou dat, které vznikají řízením topné soustavy inteligentní budovy. Cílem analýzy je nalezení vhodných metod pro modelování chování topné soustavy a detekci anomálií.

Teoretická část charakterizuje poskytnutá data a obsahuje řešerši, již známých řešení dané problematiky. Poskytnutá data jsou charakterizována jak z hlediska jich samotných, tak systému, z kterého vznikly.

Návrhová část řešení definuje postupy výběru a předzpracování poskytnutých dat a výběr vhodných metod pro modelování topné soustavy a detekci anomálií.

Implementační část řešení realizuje analýzu dat v podobě implementace programu v programovacím jazyce Python ve webové verzi pythonné konzole Jupyter Notebook.

Poslední část bakalářské práce obsahuje vyhodnocení experimentálních výsledků, závěry a doporučení.

## **1.2 Cíle práce**

1. Analýza poskytnutých dat naměřených v topných soustavách budov. Rešerše vhodných metod pro modelování chování topné soustavy a detekci anomálií.
2. Postupy výběru a předzpracování poskytnutých dat pro analýzu užitím vhodných metod pro modelování chování topné soustavy a detekci anomálií.
3. Implementace analýzy užitím vhodných metod pro modelování chování topné soustavy a detekci anomálií.
4. Provedení analýzy a vyhodnocení experimentálních výsledků, závěry a doporučení.

## 2. Analýza řešení

### 2.1 Základní pojmy

#### a) Aktuátor

Hlavice neboli akční člen, který průchodem elektrického proudu dosahuje buď aktivního či neaktivního stavu. V aktivním stavu protéká proud a hlavice je uzavřená, topné těleso netopí a teplota klesá, naopak v neaktivním stavu proud neprotéká a hlavice je otevřená, topné těleso topí a teplota stoupá. [9]

#### b) Časové řady

Posloupnost datových bodů, které jsou jednoznačně uspořádány v čase, zpravidla ve směru minulost – přítomnost. Jedná se o posloupnost diskrétních dat. Příkladem časových řad jsou naměřené teploty v určitém období, počet vyrobených výrobků v jednotlivých měsících nebo denní průtok vody potrubím. [1]

#### c) DataFrame

DataFrame je základním datovým typem knihovny Pandas. Data jsou v DataFrame v podobě řádků a jsou uspořádány ve sloupcích. DataFrame si lze představit jako obyčejnou tabulku s řádky a sloupci. [5]

#### d) DIRC

Systém DIRC (Distributed Independent Room Control) je distribuované individuální řízení prostor. Systém umožňující individuální regulaci vytápění na základě nastavené požadované teploty v budově, a to v jednotlivých zónách nezávisle. [9]

#### e) Grafana

Analytická platforma sloužící k dotazování, vizualizaci a monitorování dat. Grafana je open-source nástrojem vyvíjen společností Grafana Labs. Pro vytvoření vizualizace je nutný datový zdroj, jedním z nich může být již zmiňovaná databáze InfluxDB, další možnosti jsou

Graphite, Cloudwatch a Prometheus. K ovládání Grafany používáme ten programovací jazyk, který využívá datový zdroj. [13]

#### **f) GSM**

Síť GSM (Global System for Mobile Communications) je globální systém pro mobilní komunikaci označující nejrozšířenější datovou síť na světě. [12]

#### **g) InfluxDB**

Specificky zaměřená databáze pro data, která jsou časově označená. InfluxDB je open-source databází vyvíjená společností InfluxData. Vytvořena v programovacím jazyce Go, který umožňuje rychlé ukládání a vyhledávání dat v sériích dat v čase. InfluxDB je vhodnou volbou pro monitorování dat senzorů, síťových dat, meteorologických dat, burzovních dat a dalších typů analytických dat. [14]

#### **h) Umělý neuron**

Umělý neuron je základní funkční jednotka umělé nervové sítě. [7]

#### **i) Zóna**

Pojem zóna v oblasti vytápění znamená vymezený prostor, který je vytápěn na požadovanou teplotu. Může se jednat o konkrétní místnost v budově, ale není to pravidlem, protože zóna může např. označovat dvě a více místností nebo celé patro budovy. [10]

## **2.2 Data**

Data použitá v bakalářské práci byla poskytnutá nejmenovanou společností, která se zabývá oborem automatizace vytápěných budov. Poskytnuta data byla ve formě databáze InfluxDB získaných ze senzorů inteligentních budov, respektive z automatizovaně vytápěných budov. Data byla vizualizována prostřednictvím analytické platformy Grafana.

## 2.3 Vznik dat

### 2.3.1 Zónový systém

Význam zónového systému spočívá v individuální regulaci teploty v jednotlivých zónách. Pojem zóna je vysvětlen v kapitole č. 2. Rozmístění zón v budově, která disponuje zónovým systémem je libovolné, např. budova s 5 místnostmi může být 5-zónová, ale také nemusí, protože ne všechny místnosti musí tvořit samostatnou zónu a být individuálně regulovány teplotou. Příkladem zóny, která není tvořena pouze jednou místností jsou např. 2 místnosti, které spolu sousedí a bylo by z určitého hlediska nevhodné místnosti rozdělit do 2-zónové v tom případě, ač budova má 5 místností nebude již 5-zónová, ale pouze 4-zónová. Jakým způsobem bude v budově zónový systém navržen je pak zcela na majiteli. [9], [11]

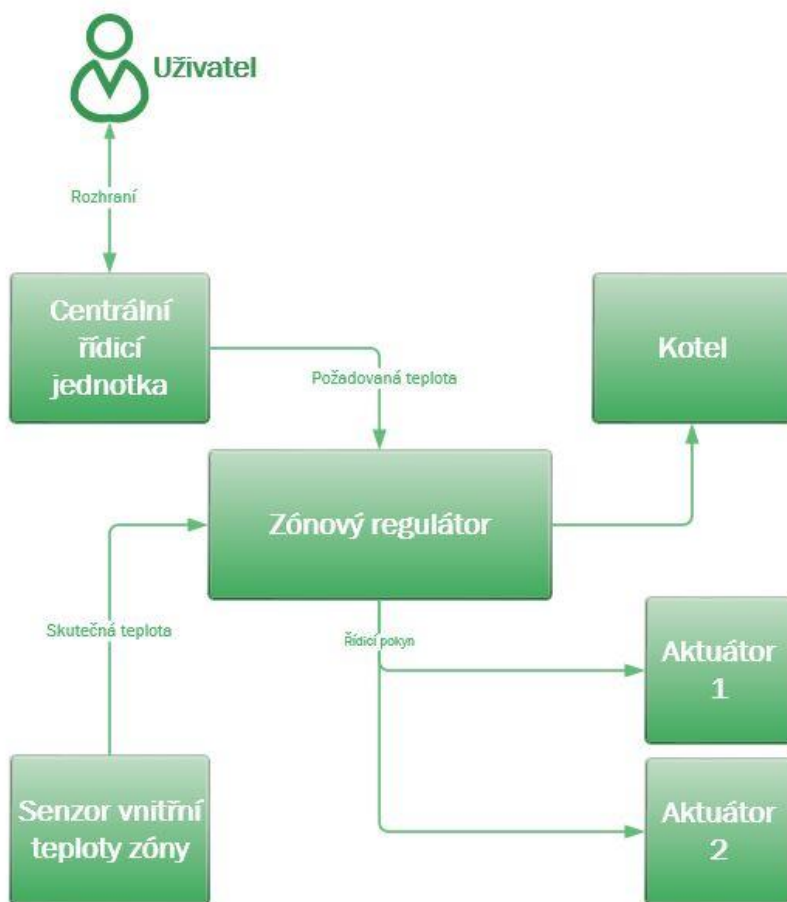
Dnešní moderní zónové systémy oproti těm starším, které využívaly kabelové propojení mezi jednotlivými zařízeními, využívají k propojení zařízení bezdrátovou technologii. Hlavní nevýhodou starších systémů s kabelovými propojeními byla instalace zónového systému do již stojící budovy. Tuto nevýhodu s příchodem moderních technologií se podařilo eliminovat za pomoci bezdrátové technologie. [2], [9]

Zónové systémy se dnes využívají jak v nekomerčních, tak i v komerčních budovách s využitím lokálního, tak i dálkového vytápění. [10]

### 2.3.2 Části zónového systému

Nejdůležitější částí zónového systému je centrální řídicí jednotka, kterou lze označit za tzv. „mozek“ celého systému. Jinak řečeno centrální řídicí jednotka slouží jako hlavní zařízení, které je nadřazené všem ostatním zařízením v zónovém systému a řídí jejich samostatnou činnost na základě nastavení uživatele. Ovládání centrální řídicí jednotky je nejčastěji realizováno prostřednictvím digitálního displeje, který slouží jako komunikační rozhraní mezi uživatelem a zónovým systémem. Veškerá nastavení zónového systému se uskutečňují v centrální řídicí jednotce. Mezi další části zónového systému patří senzory vnitřní teploty, senzory vnější teploty, termoelektrické pohony, elektronické regulátory topných těles, zónové regulátory a reléové spínací jednotky. Vývojem automatizace vytápění budov se části zónového systému rozšířily o další části, které sice systém nevyžaduje, ale zvyšují uživatelský komfort. Jedná se o senzory intenzity slunečního svitu, senzory rychlosti větru, jednotky pro ovládání žaluzií, jednotky pro ovládání otevření a zavření oken, moduly pro dálkové ovládání

prostřednictvím technologie GSM a moduly pro vizualizaci celého zónového systému v reálném čase v počítači či dokonce na Internetu. [2], [10]



Obrázek č. 1 - Schéma zónového systému

### 2.3.3 Princip zónového systému

Centrální řídicí jednotka poskytuje informaci o požadované teplotě, která byla vypočítána na základě časového harmonogramu konkrétní zóny nebo specificky nastavena uživatelem, např. uživatel v centrální řídicí jednotce nastaví, aby během noci byla topná tělese vypnuta nebo vytápěna na nejnižší možnou teplotu a stejně tak i během dne, kdy není zóna využívána, a naopak když je zóna využívána požadovanou teplotu zvýší. Informaci o požadované teplotě v dané zóně centrální řídicí jednotka odesílá do zónového regulátoru, který požadovanou teplotu porovnává se skutečnou teplotou v dané zóně. Informaci o skutečné teplotě v dané zóně zónový regulátor získává v nastavených pravidelných intervalech ze senzoru vnitřní teploty, který se nachází v dané zóně. Porovnáním požadované a skutečné teploty zónový regulátor zjišťuje, zda teplota v zóně odpovídá požadované teplotě nebo neodpovídá a na tuto skutečnost reaguje odesláním informace aktuátoru neboli hlavicí umístěné na otopném tělese.

Odeslaná informace obsahuje jeden ze dvou řídicích příkazů, který je buď otevření nebo uzavření hlavice. Otevření hlavice v dané zóně způsobí, že topné těleso začne topit a v zóně stoupá teplota. Naopak uzavření hlavice v dané zóně způsobí, že topné těleso netopí a v zóně klesá teplota. [10]

## 2.4 Struktura dat

### 2.4.1 Databáze

Data, která byla poskytnuta byla v časovém rozmezí od 2010-10-06 do 2017-10-27 a nacházela se v databázi InfluxDB. Databáze obsahuje několik měření dvou typů, jedním typem jsou data z individuální regulace teploty pod označením „<objekt>.dirc“ a druhým typem jsou data z kotelny pod označením „<objekt>.dot“.

### 2.4.2 Data z individuální regulace

Data z individuální regulace se označují „<objekt>.dirc“ a obsahují informace o čase záznamu, požadované a skutečné teplotě, příznaků aktuátorů neboli hlavice a příznak zóny. V jedné zóně je vždy minimálně 1, maximálně 4 aktuátory. K záznamu informací dochází v stanoveném pravidelném intervalu 1 hodiny. Zóny, jejichž název obsahuje označení „SCRIPT“ je nutné ignorovat. Některé záznamy je rovněž nutné ignorovat na základě některých příznaků zóny a aktuátorů. Příznaky zóny a aktuátorů fungují na principu bitových příznaků a může být platných několik příznaků současně.

Atribut	Popis	Formát
name	Název měření	object
time	Čas záznamu	int64
ta	Skutečná teplota [°C]	float64
tw	Požadovaná teplota [°C]	float64
flags	Příznak zóny	int64
a1flags	Příznak aktuátoru 1	int64
a2flags	Příznak aktuátoru 2	int64
a3flags	Příznak aktuátoru 3	int64
a4flags	Příznak aktuátoru 4	int64
zone	Název zóny	object

Tabulka č. 1 - Popis atributů dat z individuální regulace

Hodnota	Název	Popis	Vysvětlení
1	Inactive	Neaktivní	Nedochází k regulaci zóny
2	Active	Aktivní	Nedochází k regulaci zóny
4	Increase	Zvyšovat teplotu	Nedochází k regulaci zóny
8	Decrease	Snižovat teplotu	Nedochází k regulaci zóny
16	RegIncrease	Regulace zvyšovat teplotu => neaktivní	Dochází k regulaci zóny
32	RegDecrease	Regulace snižovat teplotu => aktivní	Dochází k regulaci zóny
64	RegAltIncrease	Regulace alternativní zvyšovat teplotu => neaktivní	Dochází k regulaci zóny
128	RegAltDecrease	Regulace alternativní snižovat teplotu => aktivní	Dochází k regulaci zóny
256	ConfigError	Chyba konfigurace zóny	Záznam je nutné ignorovat
512	ThermoError	Chyba snímače teploty	Záznam je nutné ignorovat
1024	FreezeError	Fungovala protimrazová ochrana	Záznam je nutné ignorovat
2048	UnaccessibleError	Nedostupná zóna	Záznam je nutné ignorovat
4096	WindowContactActive	Okenní snímač aktivní	Záznam je nutné ignorovat

Tabulka č. 2 - Popis příznaků zóny

Hodnota	Název	Popis	Vysvětlení
1	Inactive	Neaktivní aktuátor	Topné těleso topí
2	Active	Aktivní aktuátor	Topné těleso netopí
4	InactiveAntiblock	Neaktivní aktuátor – antiblok	Akční zásah, změna stavu
8	ActiveAntiblock	Aktivní aktuátor – antiblok	Akční zásah, změna stavu
16	UndercurrentError	Chyba – podproud aktuátoru	Hlavice je pravděpodobně utržená
32	FreezeError	Fungovala protimrazová ochrana	Na úrovni zónové jednotky
64	OverCurrentError	Chyba – nadproud aktuátoru	Pouze DIRC
128	Testing	Probíhalo testování aktuátoru	Pouze DIRC

Tabulka č. 3 - Popis příznaků aktuátorů



### 2.4.3 Data z kotelny

Data z kotelny se označují „<objekt>.dot“ a obsahují informace z kotelny, ohřevu teplé vody, rozdělovačů, solárů a čerpadel. K záznamu dochází v stanoveném pravidelném intervalu 15 minut.

Atribut	Popis	Formát
name	Název měření	object
time	Čas záznamu	int64
id	ID datového bodu	float64
addr	Adresa datového bodu	object
value	Hodnota datového bodu	float64

Tabulka č. 4 - Popis atributů dat z kotelny

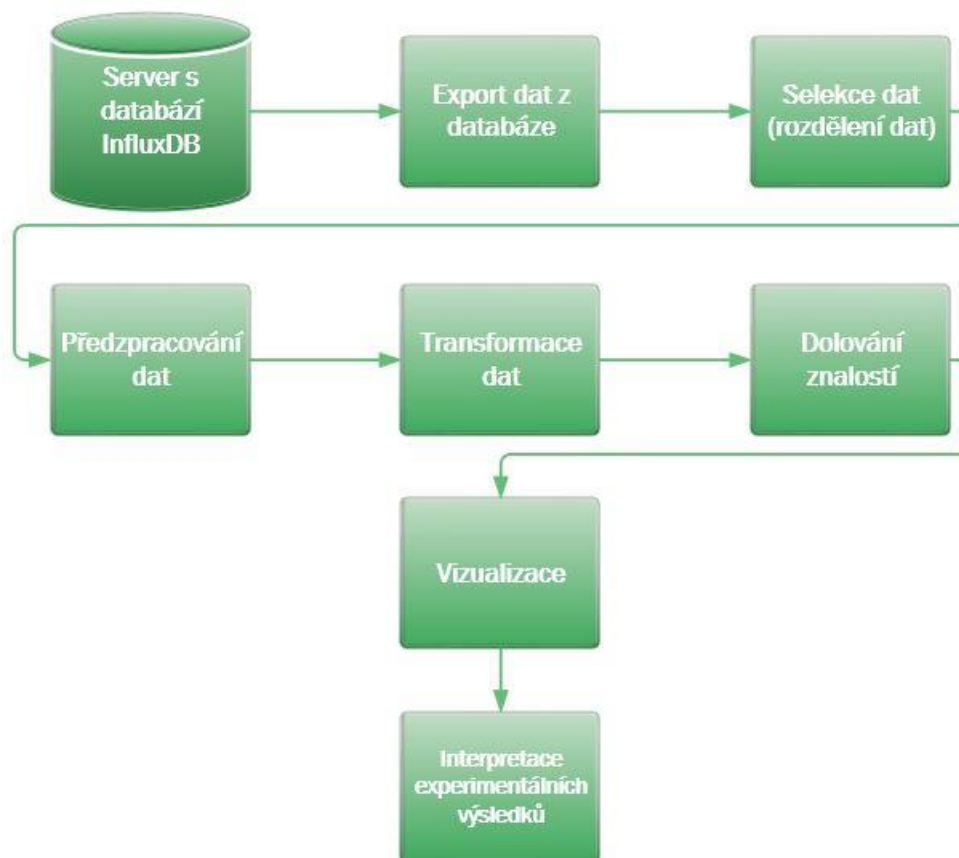
Adresa datového bodu	Popis
/K1/t0	Kotel 1 - teplota výstupní vody
/TAH4.1	Přehřátí vzduchu v kotelně
/SB4.4	STOP tlačítko stisknuto
/TAS-1.1/core	Teplota venkovního vzduchu - aktuální
/P0-2.1/core	Tlak vody - aktuální
/TNK/core	Teplota studeného vratu
/@SM	Režim systému
/Kmod/cmd	Provozní režim kotelny - cmd

Tabulka č. 5 - Popis adres datových bodů

Datové body jsou pro každou jednotlivou zónu specifické. Popis adres datových bodů (tabulka č. 5) je pouze ukázkou, neobsahuje všechny adresy datových bodů.

### 3. Návrh řešení

Návrh řešení má několik fází, které můžeme vidět na obrázku níže.



Obrázek č. 2 - Schéma návrhu řešení

Fáze interpretace experimentálních výsledků se nachází v samostatné kapitole Experimentální výsledky č. 5. [1]

#### 3.1 Export dat z databáze

##### 3.1.1 Obecný popis

Data, která byla poskytnuta k analýze jsou blíže popsána v odstavci č. 2.4. Prvním krokem analýzy dat je export dat z databáze InfluxDB, tuto fázi lze provést dvěma způsoby:

- a) InfluxDB – příkazový řádek
- b) Grafana – webová aplikace

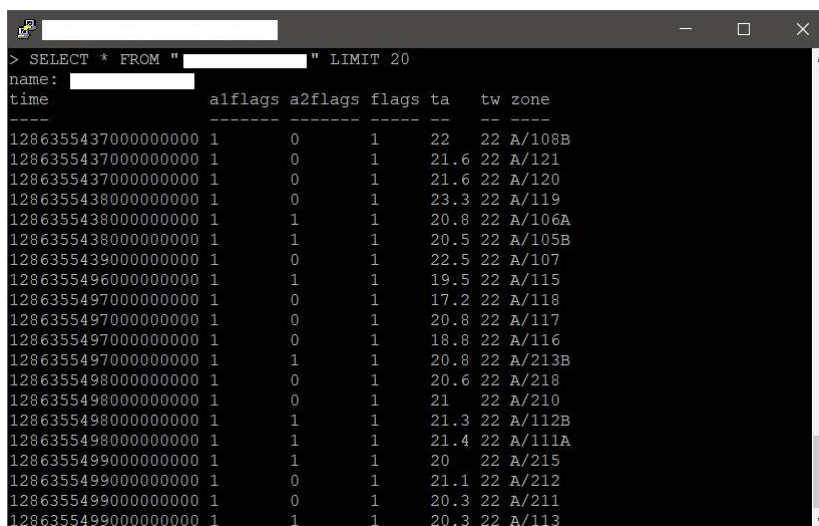
V této analýze dat byl export dat uskutečněn prostřednictvím příkazového řádku databáze InfluxDB, který umožňuje naplno využít výhod automatizovaného zpracování velkého objemu dat a následný export dat dle další potřeby. Cílem exportu dat bylo získat data jako celek a proces selekce a předzpracování dat ponechat do samostatné části analýzy dat.

### 3.1.2 Proces exportu dat

K procesu exportu dat jsem využil dvou open-source programů, jedním z nich je PuTTY a druhý je WiNSCP.

Postup:

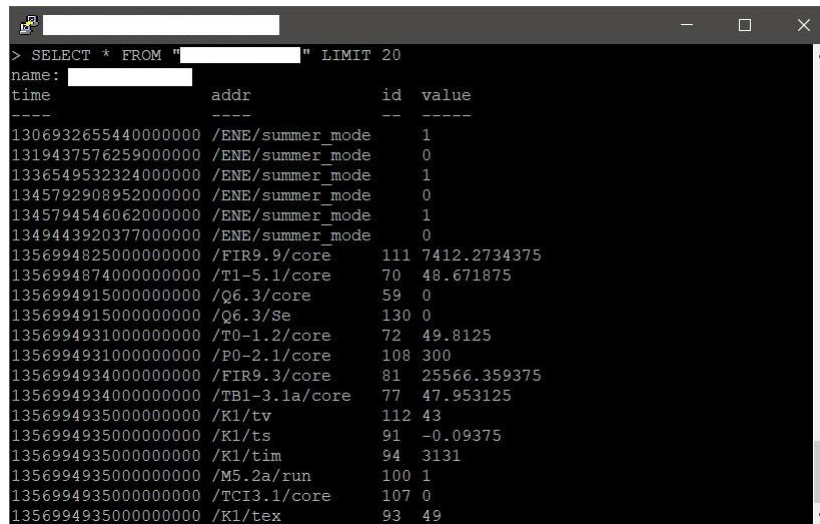
1. Připojení na server prostřednictvím programu PuTTY použitím protokolu SSH. Zahrnuje nastavení připojení, autentizace a autorizaci uživatele.
2. Spuštění databáze InfluxDB příkazem „influx“.
3. Zobrazení databází příkazem „SHOW DATABASES“.
4. Výběr databáze příkazem „USE <název\_databáze“.
5. Zobrazení tabulek či měření použité databáze příkazem „SHOW MEASUREMENTS“.
6. Zobrazení měření dat z individuální regulace, s omezením několika záznamů příkazem „SELECT \* FROM <název\_měření.dir> LIMIT 20“ (obrázek č. 3).



```
> SELECT * FROM "                  " LIMIT 20
name:                   
time                a1flags a2flags flags ta  tw zone
-----
1286355437000000000 1        0        1    22  22 A/108B
1286355437000000000 1        0        1    21.6 22 A/121
1286355437000000000 1        0        1    21.6 22 A/120
1286355438000000000 1        0        1    23.3 22 A/119
1286355438000000000 1        1        1    20.8 22 A/106A
1286355438000000000 1        1        1    20.5 22 A/105B
1286355439000000000 1        0        1    22.5 22 A/107
1286355496000000000 1        1        1    19.5 22 A/115
1286355497000000000 1        0        1    17.2 22 A/118
1286355497000000000 1        0        1    20.8 22 A/117
1286355497000000000 1        0        1    18.8 22 A/116
1286355497000000000 1        1        1    20.8 22 A/213B
1286355498000000000 1        0        1    20.6 22 A/218
1286355498000000000 1        0        1    21    22 A/210
1286355498000000000 1        1        1    21.3 22 A/112B
1286355498000000000 1        1        1    21.4 22 A/111A
1286355499000000000 1        1        1    20    22 A/215
1286355499000000000 1        0        1    21.1 22 A/212
1286355499000000000 1        0        1    20.3 22 A/211
1286355499000000000 1        1        1    20.3 22 A/113
```

Obrázek č. 3. - První náhled na data individuální regulace

7. Zobrazení měření dat z kotelny, s omezením několika záznamů příkazem „SELECT \* FROM <název\_měření.dot> LIMIT 20“ (obrázek č. 4).



```
> SELECT * FROM " " LIMIT 20
name:
time          addr          id  value
----          -
1306932655440000000 /ENE/summer_mode 1
1319437576259000000 /ENE/summer_mode 0
1336549532324000000 /ENE/summer_mode 1
1345792908952000000 /ENE/summer_mode 0
1345794546062000000 /ENE/summer_mode 1
1349443920377000000 /ENE/summer_mode 0
1356994825000000000 /FIR9.9/core      111 7412.2734375
1356994874000000000 /T1-5.1/core      70  48.671875
1356994915000000000 /Q6.3/core        59  0
1356994915000000000 /Q6.3/Se          130 0
1356994931000000000 /T0-1.2/core      72  49.8125
1356994931000000000 /P0-2.1/core      108 300
1356994934000000000 /FIR9.3/core      81  25566.359375
1356994934000000000 /TB1-3.1a/core    77  47.953125
1356994935000000000 /K1/tv            112 43
1356994935000000000 /K1/ts            91  -0.09375
1356994935000000000 /K1/tim           94  3131
1356994935000000000 /M5.2a/run        100 1
1356994935000000000 /TCI3.1/core      107 0
1356994935000000000 /K1/tex           93  49
```

Obrázek č. 4 - Prvotní náhled na data z kotelny

8. Export dat individuální regulace z databáze InfluxDB na server ve formátu csv skriptem:

```
influx -database '<název_databáze>' -execute 'SELECT * FROM
"<název_měření.dir>" -format 'csv' > /home/data_dir.csv
```

9. Export dat kotelny z databáze InfluxDB na server ve formátu csv skriptem:

```
influx -database '<název_databáze>' -execute 'SELECT * FROM
"<název_měření.dot>" -format 'csv' > /home/data_dot.csv
```

10. Připojení na server prostřednictvím programu WiNSCP použitím protokolu FTP. Zahrnující nastavení připojení, autentizaci a autorizaci uživatele.
11. Přesun vyexportovaných dat z databáze InfluxDB mezi serverem a vlastním počítačem pomocí grafického rozhraní programu WiNSCP.

Stejným způsobem z databáze InfluxDB exportují data z individuální regulace i kotelny.

## **3.2 Aplikace pro zpracování dat**

K realizaci značné části bakalářské práce, vyjma fáze Exportu dat, jsem použil webové verze pythonní konzole nazývanou Jupyter Notebook, která poskytuje velké množství knihoven, widgetu a modulů. Jupyter Notebook je součástí platformy Anaconda. Pro experimenty s neuronovými sítěmi jsem využil programu RapidMiner.

### **3.2.1 Anaconda**

Anaconda je open-source softwarový balíček poskytující ve výchozí distribuci programovací jazyk Python s možností volby verze, Jupyter Notebook, JupyterLab, Anaconda Navigator, Conda, Spyder, Glueviz a mnoho dalších běžně používaných balíčků. Anaconda Navigator je grafické uživatelské rozhraní umožňující uživateli ovládat aplikace a spravovat balíčky, bez použití příkazové řádky. [15]

### **3.2.2 Jupyter Notebook**

Jupyter Notebook je vývojově integrované webové prostředí zejména pro programovací jazyk Python, Julia a R. Instalací balíčku, lze používat i jiné programovací jazyky. Hlavní důvodem, proč jsem zvolil k analýze dat Jupyter Notebook je rozdíl od klasické konzole. Zadané příkazy a jejich výstupy jsou ve webovém rozhraní ukládány a je tedy možné se k nim vracet a dále je zpracovávat. Velkou výhodou je značkovací jazyk Markdown, který umožňuje psát text mezi příkazy a doplňovat tak celý pokus poznámkami. Kromě možnosti spouštět kód a psát text je možná i vizualizace prostřednictvím grafů, které se zobrazí rovnou na stránce ve webovém prostředí. Jupyter Notebook je uzpůsobený nejen statickým procesům, ale i dynamickým procesům, čehož lze využít např. u grafů a různých widgetů. Jednotlivé projekty tzv. „notebooky“ jsou ukládány jako textové soubory ve formátu .ipynb, což umožní jednoduché sdílení a spouštění projektů. Další možností, jak sdílet projekt je využívání určeného cloudu pro Jupyter Notebook. Případná možnost, ale ne už tak flexibilní ve všech ohledech je vygenerování z notebooku statickou HTML stránku. HTML stránka umožňuje pouze náhled do projektu a jakákoliv manipulace s projektem není možná. [16]

### 3.2.3 RapidMiner

RapidMiner je open-source aplikace pro potřeby dolování z dat. Poskytuje integrované prostředí pro selekci dat, předzpracování dat, transformaci dat, učení stroje, hluboké učení, dolování z dat a prediktivní analýzy. Výhodou RapidMineru je jednoduché a intuitivní grafické rozhraní, velká komunita aktivních uživatelů, rozsáhlý a flexibilní systém. [17]

## 3.3 Selektce dat

### 3.3.1 Obecný popis

Z důvodu velkého objemu dat, který způsobil nepřehlednost, a ne zcela dostatečné očištění dat od neplatných a chybných hodnot, detekci a odstranění odlehlých hodnot, byla data rozdělena, dle jednotlivých zón a každá jednotlivá zóna zpracována samostatně. Po zpracování jednotlivých zón a přípravě k dalšímu zpracování dle metody zpracování, data spojím v požadovaný celek a nechám podrobit dané analýze dat.

Z měření <název\_měření.dirc> obsahující data z individuální regulace jednotlivých zón pocházející z databáze <název\_databáze> vzniklo celkem 58 zón ve formátu csv.

### 3.3.2 Načtení dat

Knihovnu Pandas jsem využil k načtení dat (obrázek č. 5) z formátu csv do tzv. „DataFramu“, který umožní další zpracování a export po selekci dat prostřednictvím widgetu Qgrid.

```
# import modulu
import pandas as pd
import qgrid as qg

# cesta k souboru s daty
path = "C://Users//Zemi4//Desktop//Data//data_dirc.csv"

# nacteni dat do DataFramu
data = pd.read_csv(path, delimiter = ",")
```

Obrázek č. 5 - Kód načtení dat

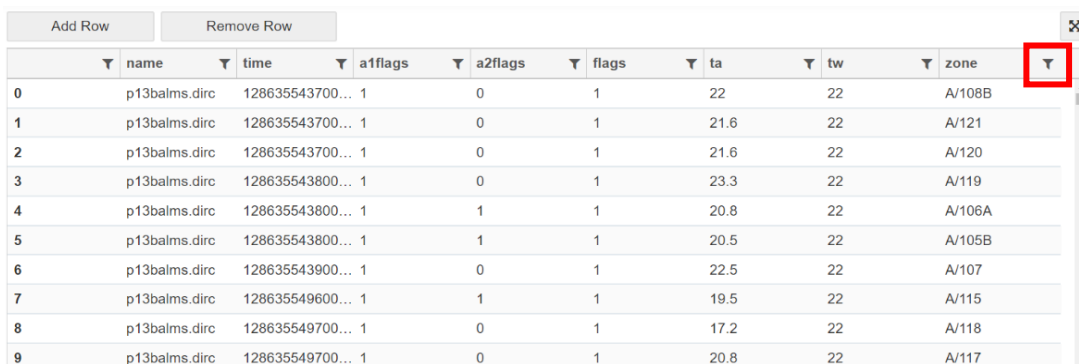
### 3.3.3 Rozdělení dat

Widget Qgrid (obrázek č. 6) poskytuje interaktivní prostředí, které poskytne vizualizaci dat a selekci dat na základě filtrů. Velkou výhodou je, že widget Qgrid pracuje dynamicky a nemusí se opětovně spouštět kód. Využitím filtrů jsem provedl rozdělení dat dle jednotlivých zón (obrázek č. 7 a 8). [18]

```
# vizualizuje DataFrame jako interaktivni QgridWidget
data_widget = qg.show_grid(data, show_toolbar = True)

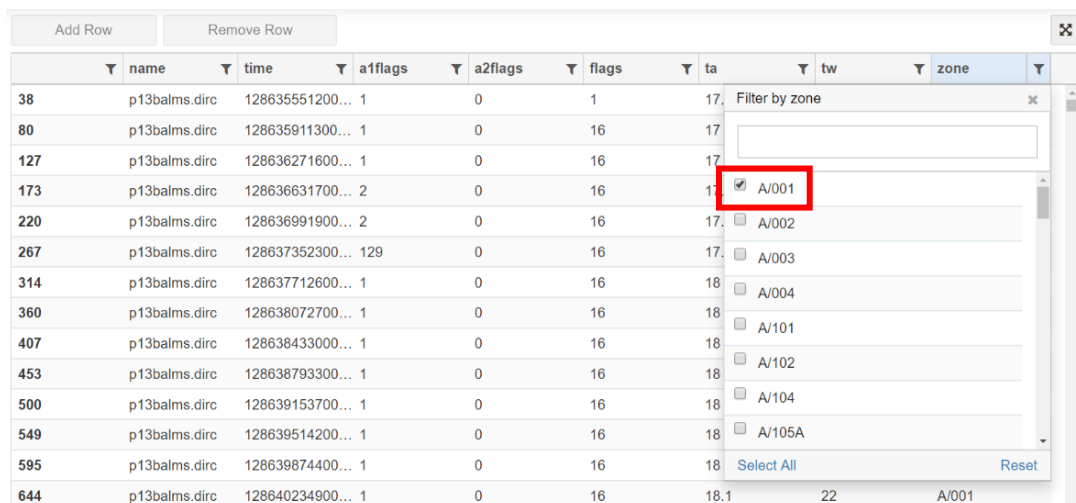
# zobrazeni interaktivniho QgridWidget
data_widget
```

Obrázek č. 6 - Kód rozdělení dat I.



	name	time	a1flags	a2flags	flags	ta	tw	zone
0	p13balms.dirc	128635543700...	1	0	1	22	22	A/108B
1	p13balms.dirc	128635543700...	1	0	1	21.6	22	A/121
2	p13balms.dirc	128635543700...	1	0	1	21.6	22	A/120
3	p13balms.dirc	128635543800...	1	0	1	23.3	22	A/119
4	p13balms.dirc	128635543800...	1	1	1	20.8	22	A/106A
5	p13balms.dirc	128635543800...	1	1	1	20.5	22	A/105B
6	p13balms.dirc	128635543900...	1	0	1	22.5	22	A/107
7	p13balms.dirc	128635549600...	1	1	1	19.5	22	A/115
8	p13balms.dirc	128635549700...	1	0	1	17.2	22	A/118
9	p13balms.dirc	128635549700...	1	0	1	20.8	22	A/117

Obrázek č. 7 - Kód rozdělení dat II.



	name	time	a1flags	a2flags	flags	ta	tw	zone
38	p13balms.dirc	128635551200...	1	0	1	17.		Filter by zone
80	p13balms.dirc	128635911300...	1	0	16	17		
127	p13balms.dirc	128636271600...	1	0	16	17		
173	p13balms.dirc	128636631700...	2	0	16	17		<input checked="" type="checkbox"/> A/001
220	p13balms.dirc	128636991900...	2	0	16	17.		<input type="checkbox"/> A/002
267	p13balms.dirc	128637352300...	129	0	16	17.		<input type="checkbox"/> A/003
314	p13balms.dirc	128637712600...	1	0	16	18		<input type="checkbox"/> A/004
360	p13balms.dirc	128638072700...	1	0	16	18		<input type="checkbox"/> A/101
407	p13balms.dirc	128638433000...	1	0	16	18		<input type="checkbox"/> A/102
453	p13balms.dirc	128638793300...	1	0	16	18		<input type="checkbox"/> A/104
500	p13balms.dirc	128639153700...	1	0	16	18		<input type="checkbox"/> A/105A
549	p13balms.dirc	128639514200...	1	0	16	18		
595	p13balms.dirc	128639874400...	1	0	16	18		
644	p13balms.dirc	128640234900...	1	0	16	18.1	22	A/001

Obrázek č. 8 - Kód rozdělení dat III.

### 3.3.4 Uložení dat

K uložení dat jsem opět využil knihovny Pandas (obrázek č. 9), která uloží DataFrame do csv souboru. Data jsou rozdělena do několika csv souborů, dle jednotlivých zón a v další fázi práce se věnuji předzpracování dat pro další zpracování.

```
# ziska kopii DataFramu, která byla k vytvoreni aktualni instance QgridWidget
sel_data = data_widget.get_changed_df()

# cesta k adresari ulozeni dat
path = "C://Users//Zemi4//Desktop//A-001.csv"

# ulozeni DataFramu do formatu .csv soubrou
sel_data.to_csv(path, encoding = 'utf-8', index = False)
```

Obrázek č. 9 - Kód uložení dat

## 3.4 Předzpracování dat

### 3.4.1 Obecný popis

Předzpracování dat pro další zpracování takového rozsahu bylo velmi obtížné, způsobené z důvodů zmíněných v odstavci č. 3.3.1. Řešením bylo rozdělení celku dat dle jednotlivých zón do samostatných souborů.

### 3.4.2 Načtení a uložení dat

K načtení a uložení dat jsem využil knihoven OS a Glob (obrázek č. 10 a 11), které mi umožnily načíst soubor zóny, následně ho zpracovat a výsledek zpracování uložit do nového souboru. Tento postup se aplikuje na celý adresář se soubory a jednotlivé zóny jsou postupně předzpracovány.

```
# cesta k adresari se seoubory s nepredzpracovanymi daty
folder_path = 'C:/Users/Zemi4/Desktop/Selekce_dat/'

# cesta k adresari k ulozeni souboru s predzpracovanymi daty
output_path = 'C:/Users/Zemi4/Desktop/Predzpracovani_dat/'

# funkce k nacteni vseh souboru ve slozce k predzpracovani
for file in glob.glob(folder_path + '*.csv'):

    # nacteni obsahu souboru do DataFramu
    df = pd.read_csv(file, delimiter = ";")
```

Obrázek č. 10 - Kód načtení a uložení dat I.



```
# uložení DataFramu do .csv souboru
vysledek_df.to_csv(output_path + str(os.path.basename(file)),
                    encoding = 'utf-8', index = False)
```

Obrázek č. 11 - Kód načtení a uložení dat II.

### 3.4.3 Formát času

Formát času v poskytnutých datech je v unixovém formátu času. Tento formát času pro interpretaci nebo vizualizaci dat není praktický. Z toho důvodu jsem zvolil přeformátování tohoto atributu do formátu dle normy ISO rok, měsíc, den a hodiny, např. 2010-12-24 12:00:00. K přeformátování času jsem použil knihovnu Datetime (obrázek č. 12).

```
# prevod formatu atributu "time" do formatu string retezce
# prekopirovani do noveho vytvoreneho atributu "time_str"
df["time_str"] = [dt.datetime.strptime(d, "%d.%m.%Y %H:%M:%S") for d in df["time"]]

# vybrani z datumove casti pouze rok, mesic a den
df["time_str"] = [d.date() for d in df["time_str"]]

# prevod formatu atributu "time_str" do formatu datetime
df["time_str"] = pd.to_datetime(df["time_str"])

# prekopirovani obsahu atributu "time_str" do noveho atributu "time_uprav"
df["time_uprav"] = df["time_str"]
```

Obrázek č. 12 - Kód formát času

### 3.4.4 Funkce pro analýzu souboru

Analýza souboru zahrnuje několik kroků a pro zjednodušení rozeberu každý krok zvlášť.

#### 3.4.4.1 Selektce atributů

Pro další zpracování dat jsou některé atributy zbytečné, z toho důvodu jsem provedl selekci těchto atributů (obrázek č. 13).

```
# vyber atributu k analyze
new_df = df.loc[df["time_str"] == pozadovane_data, ["ta", "tw", "zone", "a1flags",
                                                    "a2flags", "time_uprav", "flags"]]
```

Obrázek č. 13 - Kód selektce atributů

Nepotřebné atributy	Požadované atributy
name	ta, tw, zone, a1flags, a2flags, flags, time_uprav

Tabulka č. 6 - Selekce atributů

### 3.4.4.2 Počet měření

V ideálním stavu měření individuální regulace v jednotlivých zónách za jeden den proběhne přesně 24 měření čili každou hodinu jedno měření. Tento ideální stav nebyl vždy naplněn, což mohlo být způsobeno chybou zařízení, výpadkem elektřiny, chybou obsluhy atd. a stávalo se, že počet měření neodpovídal 24, tudíž počet měření bylo více či méně. Z toho důvodu bylo nutné u každého dnu jednotlivých zón provést kontrolu ideálního stavu, zda naplňuje počet 24 měření za jeden den a pokud tomu tak nebylo, buď tento počet doplnit nebo odstranit měření, která byla navíc. K úpravě počtu měření jsem použil knihovny NumPy a Pandas (obrázek č. 14).

```
# vypocet poctu mereni za dany den
counter = new_df.ta.count()

# pokud je pocet mereni za dany den mensi nez 24, dojde k doplneni mereni
if counter < 24:
    for i in range(counter, 24):
        new_df.loc[i] = [Nan for n in range(7)]
        new_df["ta"] = new_df.ta.fillna(0)
        new_df["tw"] = new_df.tw.fillna(0)
        new_df["zone"] = new_df.zone.fillna(0)
        new_df["a1flags"] = new_df.a1flags.fillna(0)
        new_df["a2flags"] = new_df.a2flags.fillna(0)
        new_df["flags"] = new_df.a2flags.fillna(0)
        new_df["time_uprav"] = new_df.time_uprav.fillna(new_df.time_uprav.min())

# pokud je pocet mereni za dany den vetsi nez 24, dojde k odstraneni poslednich mereni
elif counter > 24:
    counter_list = list(range(24, counter))
    new_df = new_df.drop(new_df.index[counter_list])
```

Obrázek č. 14 - Kód počet měření

### 3.4.4.3 Čas měření

Z prvotních vizualizací dat vyplynulo, že data nejsou jednotné v čase záznamu měření, protože časy záznamů z individuální regulace jednotlivých zón, byly sice měřeny v intervalech jedné hodiny, ale pro každou jednotlivou zónu v jiný čas. Tento čas se lišil v několika minutách až desítkách minut v rozsahu od 45. minuty do 15. minuty následující hodiny. Z toho důvodu bylo potřeba sjednotit všechny časy měření provedené v jednom měření na jednotný čas, který jsem stanovil jako každou celou hodinu, např. 12:00, 22:00 nebo 00:00. K úpravě času měření jsem použil knihovnu Pandas (obrázek č. 15 a 16).

```

# vezme datumovou cast z "time_uprav" a prida hodinovou cast
new_df["time_oprava"] = [dt.datetime.combine(d.date(),dt.time(1, 0))
                        for d in new_df["time_uprav"]]

# vytvoreni prazdneho listu pro hodinovy rozsah
s = 0
cas_list = []

```

Obrázek č. 15 - Kód čas měření I.

```

# naplneni listu hodnotami datum + spravna hodinova cast
for d in new_df["time_oprava"]:
    d = d + dt.timedelta(hours = s)
    cas_list.append(d)
    s = s + 1

# prevod listu na serii
se = pd.Series(cas_list)

# prepsani atributu "time_oprava" listem hodnot od 01:00 do 00:00
new_df['time_oprava'] = se.values

```

Obrázek č. 16 - Kód čas měření II.

#### 3.4.4.4 Validace měření

Doplněním chybějících záznamů měření bylo nutné provést validaci měření, která nám říká, zda se měření uskutečnilo nebo neuskutečnilo.

```

# provedeni validace mereni
# pokud bylo mereni doplnene analyzou, jedna se o neplatne mereni
new_df['val'] = (new_df['ta'] != 0) & (new_df['tw'] != 0) & (new_df['zone'] != 0)

```

Obrázek č. 17 - Kód validace měření

#### 3.4.4.5 Přejmenování atributů

Některé názvy atributů nebyly dostatečně jasné a byly ve zkrácené formě, což mohl vést k nepřehlednosti. Z toho důvodu jsem sloupce přejmenoval do lépe srozumitelné formy.

```

# prejmenovani sloupcu
new_df.rename(columns = {"ta" : "sku_tep", "tw" : "poz_tep", "a1flags" : "a1_pri",
                        "a2flags" : "a2_pri", "time_oprava" : "cas", "zone" : "zona",
                        "flags" : "zona_pri"}, inplace = True)

```

Obrázek č. 18 - Kód přejmenování atributů

Původní název atributu	Nový název atributu
time	cas
ta	sku_tep
tw	poz_tep
flags	zona_pri
a1flags	a1_pri
a2flags	a2_pri
zone	zona

Tabulka č. 7 - Přejmenování atributů

### 3.4.5 Funkce pro všechny dny

Pro zpracování všech dní měření a zkontrolování počtu měření (odstavec č. 3.4.4.2 a 3.4.4.3) za daný den bylo nutné vytvořit funkci, která zavolá funkci analýzy souboru na všechny dny měření.

#### 3.4.5.1 Časové rozmezí

V této části je nutné ručně zadat časové rozmezí od kdy do kdy bude volána funkce (obrázek č. 19). Poskytnutá data byla v časovém rozmezí od 2010-06-10 do 2017-27-10.

```
# uložení začátku časového rozsahu do DataFramu
start = dt.datetime(2010, 10, 6)

# uložení konce časového rozsahu do DataFramu
end = dt.datetime(2017, 10, 27)
```

Obrázek č. 19 - Kód časové rozmezí I.

Dále jsem vytvořil prázdný list, který jsem na základě časového rozsahu naplnil.

```
# vytvoření prázdného listu pro datumový rozsah
date_range = []

# vyplnění listu časového rozsahu, dle DataFramu 'start' a 'end'
date_range = [start + dt.timedelta(days = x)
               for x in range(0, (end - start).days)]
```

Obrázek č. 20 – Kód časové rozmezí II.

### 3.4.5.2 Funkce časový rozsah

Posledním krokem bylo vytvořit funkci, která zavolá na daný časový rozsah funkci analýzy souboru a uloží výsledek do DataFramu.

```
# zavolani funkce pro vsechny dny a ulozeni do DataFramu
for d in date_range:
    pom = analiza(d)
    vysledek_df = vysledek_df.append(pom, ignore_index = True)
```

Obrázek č. 21 - Kód funkce časový rozsah

### 3.4.5.3 Odstranění atributů

Atributy, které byly nutné vytvořit k předzpracování dat a nemají pro další zpracování význam byly odstraněny.

```
# odstraneni atributu, ktere nemaji dalsi vyznam pro zpracovani
vysledek_df.pop('time_uprav')
```

Obrázek č. 22 - Kód odstranění atributů

## 3.4.6 Závěr předzpracování dat

Následující předzpracování dat je základní a pro další zpracování je nutná fáze transformace dat, která přizpůsobí data podle potřeb analytické metody.

## 3.5 Transformace dat

### 3.5.1 Obecný popis

Transformace dat podle potřeb použitých analytických metod. Může zahrnovat selekci atributů, vážení atributů, normalizaci atributů, funkční transformaci a doplňování dat atributy s výpočetními hodnotami.

### 3.5.2 Doplnění atributů

Vizualizace dat prostřednictvím teplotní mapy vyžadovala doplnění dat o další atributy, které původní data neobsahují a vizualizace je vyžaduje. Atributy, které jsem doplnil jsou dva. Prvním atributem je rozdíl skutečné a požadované teploty (obrázek č. 23), druhým atributem je derivace skutečné teploty (obrázek č. 24) na základě, které se ve fázi dolování dat vypočítá

korelace mezi jednotlivými zónami. Doplnění těchto dvou atributů jsem zahrnul již do fáze předzpracování dat.

```
# rozdíl atributu "tw" a "ta" neboli požadované a skutečné teploty
new_df['roz'] = new_df['ta'] - new_df['tw']
```

Obrázek č. 23 - Kód doplnění atributů I.

```
# derivace atributu "ta" neboli skutečné teploty
new_df['der'] = new_df['ta'].diff()
```

Obrázek č. 24 - Kód doplnění atributů II.

### 3.5.3 Selektce hodnot atributů

Hodnoty atributů příznaků dvou aktuátorů a zóny, které vizualizují prostřednictvím spojnicového grafu vyžadovaly selekci hodnot atributů, protože obsahují několik druhů příznaků. Příznaky aktuátorů jsem rozdělil na aktivní a neaktivní aktuátor. Příznaky zóny jsem rozdělil na regulující a neregulující zónu. Selektci hodnot atributů jsem provedl přemapováním hodnot atributů (obrázek č. 25), které jsem v tabulkách č. 8 a 9 níže definoval.

```
# hodnoty k zmapování
map_a = {2 : 1, 8 : 1, 1 : 0, 4 : 0, 16 : 0,
         32 : 0, 64 : 0, 128 : 0, 129 : 0, 130 : 0}

# zmapuje atribut "a1_pri" dle DataFrame 'map_a'
df['a1_pri'] = df['a1_pri'].map(map_a)
```

Obrázek č. 25 – Kód zmapování hodnot atributů

Stejným způsobem pouze s jinými hodnotami jsem nahradil hodnoty atributu příznaku zóny.

#### a) Příznaky aktuátorů

Příznak aktuátor	Nahrazení příznaky	Vysvětlení
2, 8	1	Aktivní aktuátor
1, 4, 16, 32, 64, 128	0	Neaktivní aktuátor

Tabulka č. 8 – Příznaky aktuátorů

## b) Příznaky zóny

Příznak zóny	Nahrazení příznaky	Vysvětlení
16, 64	2	Regulující zóna – zvyšování teploty
32, 128	1	Regulující zóna – snižování teploty
1, 2, 4, 8, 256, 512, 1024, 2048, 4096	0	Neregulující zóna

Tabulka č. 9 - Příznaky zóny

## 3.5.4 Selektce atributů

Selektci atributů jsem provedl na úrovni souborů tak, že jsem vytvořil dvě funkce, které mi vždy na základě požadavku analytické metody spojily jednotlivé zóny. Rozlišuji dva druhy spojení, jedním je spojení jednotlivých zón do celku a druhým je na základě určitého atributu. Základ obou funkcí je stejný, rozdíl je v způsobu spojení a v počtu atributů.

### a) Spojení do celku

```
# spojení DataFramu podle atributu "cas", od shora dolu a uložení do DataFramu  
fr = pd.concat([df1, df2], axis = 0, ignore_index = True)
```

Obrázek č. 26 - Spojení do celku I.

	a1_pri	a2_pri	cas	der	poz_tep	roz	sku_tep	val	zona	zona_pri
0	1.0	0.0	2010-10-06 01:00:00	NaN	22.0	-4.7	17.3	True	A/001	0.0
1	1.0	0.0	2010-10-06 02:00:00	-0.3	22.0	-5.0	17.0	True	A/001	0.0
2	1.0	0.0	2010-10-06 03:00:00	0.0	22.0	-5.0	17.0	True	A/001	0.0
3	2.0	0.0	2010-10-06 04:00:00	0.9	22.0	-4.1	17.9	True	A/001	0.0
4	2.0	0.0	2010-10-06 05:00:00	0.0	22.0	-4.1	17.9	True	A/001	0.0

Obrázek č. 27 - Spojení do celku II.

## b) Spojení dle atributu

```
# spojení DataFramu podle atributu "cas", od leva do prava a uložení do DataFramu  
fr = pd.concat([df1.set_index('cas'), df3.set_index('cas')], axis = 1, join = 'inner').reset_index()
```

Obrázek č. 28 - Spojení dle atributu I.

	cas	A-001	A-002	A-003	A-004	A-101	A-102	A-104	A-106A	A-106B	...	A-214B	A-215	A-217	A-218	A-219	A-221	A-222	A-223	A-225	A-226
0	2010-10-07 01:00:00	18.0	17.5	18.2	18.3	18.3	18.8	18.5	20.0	20.2	...	20.5	19.8	21.0	20.5	19.8	22.8	20.3	21.0	20.8	19.8
1	2010-10-07 02:00:00	18.0	17.4	18.2	18.3	18.2	18.7	18.5	19.9	20.1	...	20.4	19.8	21.0	20.5	19.7	22.8	20.5	21.1	21.0	20.1
2	2010-10-07 03:00:00	18.0	17.5	18.1	18.3	18.3	18.7	18.5	19.9	20.1	...	20.4	19.8	21.0	20.5	19.7	22.7	20.6	21.8	21.0	20.3
3	2010-10-07 04:00:00	18.0	17.4	18.1	18.2	18.3	18.7	18.5	19.9	20.1	...	20.3	19.8	21.0	20.4	19.6	22.6	20.7	21.3	21.1	20.3
4	2010-10-07 05:00:00	18.0	17.4	18.1	18.3	18.5	18.7	18.7	20.0	20.2	...	20.4	19.8	20.9	20.5	19.7	22.7	20.7	21.3	21.1	20.2

Obrázek č. 29 - Spojení dle atributu II.

## 3.6 Dolování znalostí

### 3.6.1 Obecný popis

Dolování znalostí využívá metody umělé inteligence, metody založené na strojovém učení, metod shlukové analýzy, metod modelování a automatické tvorby modelů a širokou škálu klasifikátorů.

### 3.6.2 Korelace

Korelace je statistické měřítko, které udává míru vzájemného vztahu dvou nebo více proměnných. Korelace mezi proměnnými vyjadřuje určitou závislost, ale nelze z této závislosti určit, že jedna proměnná způsobí nebo ovlivní změnu v jiné proměnné, protože může existovat další faktor, který ovlivňuje obě proměnné. [8] [22]

#### 3.6.2.1 Korelační koeficient

Míru korelace definuje korelační koeficient, který označujeme jako statistickou míru. Vztah mezi proměnnými může být pozitivní, nebo negativní. Hodnota korelačního koeficientu se vyjadřuje v rozmezí od -1 do +1. [21]

#### 3.6.2.2 Pozitivní korelace

Pozitivní korelace je vztah mezi dvěma proměnnými, který označuje míru, do které se proměnné společně zvyšují nebo snižují. Dokonalá pozitivní korelace, vyjádřena hodnotou



korelačního koeficientu jako +1, značí zcela přímou závislost mezi proměnnými neboli zvýšení či snížení hodnot v jedné proměnné vždy odpovídá zvýšení či snížení hodnot druhé proměnné. Jestliže hodnoty dvou proměnných někdy zcela neodpovídají trendu zvyšování či snižování, tak je hodnota korelačního koeficientu je vyjádřena větší jako 0, ale menší než +1. Příkladem pozitivní korelace je např. chudoba a četnost zločinu, vzdělání a výše příjmu, cvičení a růst svalové hmoty, učení a zlepšení prospěchu. [23]

### 3.6.2.3 Negativní korelace

Negativní korelace je vztah mezi dvěma proměnnými, který označuje míru, kdy jedna proměnná se zvyšuje a druhá proměnná klesá. Dokonalá negativní korelace vyjádřena hodnotou korelačního koeficientu jako -1, značí zcela nepřímou závislost neboli zvýšení hodnot v jedné proměnné zcela odpovídá snížení hodnot v druhé proměnné a naopak. Pokud negativní korelace není dokonalá hodnota korelačního koeficientu je vyjádřena menší než 0, ale více než -1. Příkladem negativní korelace je např. náboženství a nehodovost, počasí a výhra v loterii, vzdělání a zbožnost náboženství. [24]

### 3.6.2.4 Nulová korelace

Hodnota korelačního koeficientu může nabývat hodnoty nula, která vyjadřuje nedostatek korelace. Chybí sklon hodnot proměnných k pozitivní nebo negativní korelaci, pak mezi proměnnými není žádná zjistitelná závislost. [24]

### 3.6.2.5 Pearsonův korelační koeficient

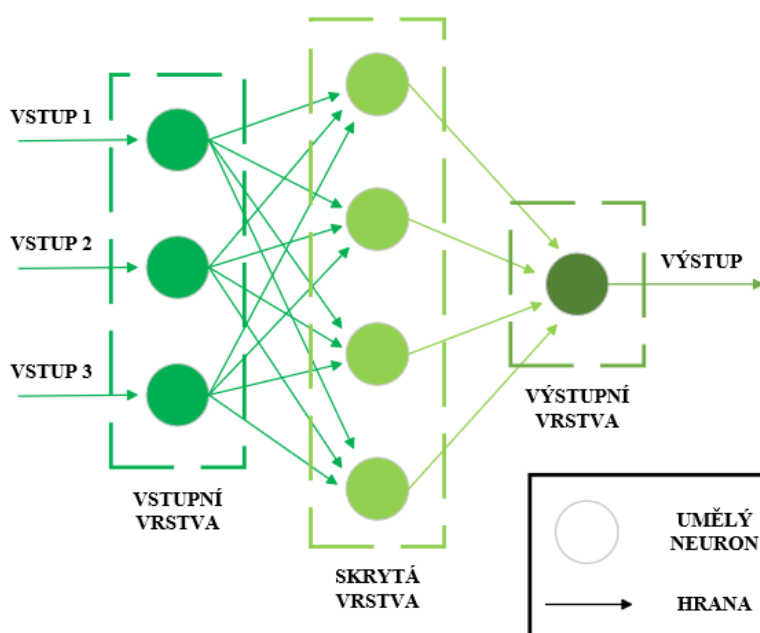
Jeden z nejznámějších korelačních koeficientů je Pearsonův, někdy také nazýván jako párový, charakterizuje lineární závislost mezi dvěma proměnnými. Pro nelineární závislosti je Pearsonův korelační koeficient nepoužitelný. Výpočet dle vzorce:

$$r_{x,y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)} - \sqrt{E(Y^2) - E^2(Y)}}$$

Výpočtem zjistíme míru závislosti dvou proměnných. Hodnota může nabývat od -1 do 1. Pokud je hodnota koeficientu rovna 0, závislost mezi proměnnými není. [8], [20]

### 3.6.3 Umělé neuronové sítě

ANN (Artificial Neural Networks) v překladu umělá neuronová síť je tvořena matematickými modely implementovány výpočetními systémy, vycházející z biologického nervového systému živého organismu. Umělá neuronová síť se skládá z jednotlivých uspořádaných umělých neuronů do struktury, která tvoří topologii umělé neuronové sítě. Jednotlivé umělé neurony navzájem spolu komunikují prostřednictvím hran. Hrany umožňují umělým neuronům řešit konkrétní problém společně. Druhy umělých neuronových sítí rozeznáváme na základě typu umělého neuronu, topologickým uspořádáním umělých neuronů a přístupem k učení sítě. Umělé neuronové sítě se učí stejným způsobem jako živý organismus příkladem. V dnešní době se umělé neuronové sítě hojně využívají např. na rozpoznání vzorů psaného písma, řeči, obrazců, zvuku nebo klasifikaci dat, prostřednictvím procesu učení. [6], [25]



Obrázek č. 30 - Dopředná neuronová síť

#### 3.6.3.1 Princip neuronové sítě

Princip umělé neuronové sítě vidíme na obrázku č. 30, jedná se o nejprimitivnější umělou neuronovou síť, která se nazývá jako dopředná neuronová síť. Z obrázku můžeme vidět, že dopředná neuronová síť šíří signál jednosměrně, a to od vstupu sítě k výstupu sítě. Neurony dopředné neuronové sítě jsou seřazeny do tří vrstev vstupní, skryté a výstupní. Hrany jsou orientované pouze na jednu stranu, a to směrem od vstupu sítě k výstupu sítě, toto řešení není pravidlem a většina neuronových sítí má propojení i v opačném směru. Hrany mezi neurony

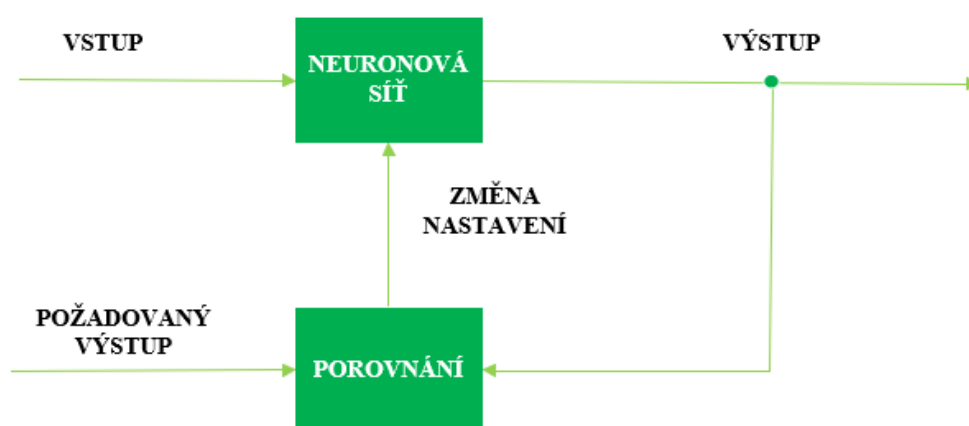
zprostředkovávají šíření signálu a každá hrana má svou váhu, která udává sílu šířícího signálu. [6], [25]

### 3.6.3.2 Fáze neuronové sítě

Neuronová síť má dvě fáze, které se nazývají adaptační a aktivní. V adaptační fázi, která probíhá jako první dochází k nastavování hran a vah jednotlivých neuronů tak, aby neuronová síť naplnila požadované cíle. Nastavení vah a hran neuronové sítě je možné dosáhnout dvěma způsoby učením a výpočtem. Možnost nastavení výpočtem není možné u všech typů neuronových sítí. Možnost nastavení učení má dvě různé cesty učení s učitelem a bez učitele. [6], [25]

### 3.6.3.3 Učení s učitelem

V případě procesu učení s učitelem je k dispozici neuronové síti nejen vstup, ale i učitel. Na vstupu neuronové sítě je tedy vstup a požadovaný výstup. Neuronová síť pro každý vstup vytvoří výstup a ten porovnává s požadovaným výstupem, který je k dispozici. Na základě porovnání a zjištění odchylky proces učení upraví, zejména nastavení vah a prahů neuronové sítě tak, aby odchylka mezi výstupem a požadovaným výstupem byla minimální nebo žádná. [7]



Obrázek č. 31 - Učení s učitelem

### 3.6.3.4 Učení bez učitele

V případě procesu učení bez učitele je k dispozici neuronové síti pouze vstup, učitel neboli požadovaný výstup chybí. Z toho důvodů neuronové síti nezbyvá nic jiného než hledat

zákonitosti ve vstupních datech a na základě toho nastavit váhy a hrany. Cílovým stavem učení bez učitele je podobným vstupním datům generovat podobná výstupní data. [7]



Obrázek č. 32 - Učení bez učitele

## 3.7 Vizualizace

### 3.7.1 Obecný popis

Vizualizace poskytuje interpretaci výsledků v podobě teplotní mapy a spojnicového grafu.

### 3.7.2 Teplotní mapa

Teplotní mapa je grafická vizualizace dat, kde každá hodnota proměnné je znázorněná určitou barvou, která je definována měřítkem v barevném spektru. Implementace teplotní mapy prostřednictvím knihovny Seaborn. Výpočet korelace umožňuje knihovna Pandas prostřednictvím funkce `corr()`, která realizuje Pearsonův korelační koeficient. [19]

#### a) Teplotní mapa – rozdíl skutečné a požadované teploty

```
# korelace derivace skutečne teploty
corr = df1.corr()

# vytvoreni zakladu grafu
plt.figure(figsize=(10,10))

# vytvoreni teplotni mapy
sns.heatmap(corr, linewidth = 0.5, annot=True, center=0, vmin = -1, vmax = 1)
```

Obrázek č. 33 – Implementace teplotní mapy rozdíl

## b) Teplotní mapa – korelace jednotlivých zón

```
# selekce atributu pro vizualizaci teplotní mapy
df2 = df1[["cas", "zona", "roz"]].copy()

# vytvoreni kontigencni tabulky
df_plt = df2.pivot_table(index = "zona", columns = "cas", values = "roz")

# vyvoreni zakladu grafu
plt.figure(figsize = (30, 30), dpi = 500)

#vytvoreni teplotni mapy
sns.heatmap(df_plt, center = 0, vmin=-10, vmax=13)
```

Obrázek č. 34 – Implementace teplotní mapa korelace

### 3.7.3 Spojnicový graf

Spojnicový graf je jedním z nejprimitivnějších typů grafů. Nejčastější využití spojnicového grafu je vizualizace souvisejících dat neboli hodnot v čase, např. teploty, příznaků a hmotnosti. Jednotlivé hodnoty v grafu jsou spojeny čarou. Implementace spojnicového grafu prostřednictvím knihovny Matplotlib. [1]

```
# vytvoreni kontigencni tabulky
df_plt = df2.pivot_table(index = "cas", columns = "zona")

# vytvoreni zakladu grafu
fig = plt.figure(figsize = (50, 50))

# vytvoreni grafu
df_plt.plot()
```

Obrázek č. 35 - Implementace spojnicový graf

## 4. Implementace řešení

Implementace analýzy dat je v programovacím jazyce Python ve webové verzi pythonní konzole Jupyter Notebook (odstavec č. 3.2.2), která je součástí platformy Anaconda (odstavec č. 3.2.1). Implementace jednotlivých částí analýzy dat je situována do tzv. „notebooků“. Kód v takovém notebooku je uspořádán do tzv. „buněk“. Buňky jsou spravovány jako samostatné celky, což umožňuje každé buňce mít vlastní vstup a výstup nezávisle na ostatních buňkách. Jupyter Notebook kombinuje programovací jazyk s dokumentačním jazykem. To mi umožnilo vytvořit interaktivní programy, které je možné jednoduše spravovat a sdílet. Implementační řešení analýzy dat je tvořeno 5 notebooky, v závorce uvedené názvy souboru notebooku:

- Selekcce dat (selekce\_dat.ipynb)
- Předzpracování dat (predzpra\_dat.ipynb)
- Transformace dat (transfor\_dat.ipynb)
- Přetápění a promrzání zón (roz\_zon.ipynb)
- Korelace zón (kor\_zon.ipynb)
- Chování zóny (chovani\_zon.ipynb)

Řešení posledního experimentu je implementováno prostřednictvím programu RapidMiner v podobě procesů, v závorce uvedené názvy procesů:

- Predikce příznaku zóny (neuron\_sit , appl\_model)

Knihovny použité v implementaci řešení jsou Pandas, NumPy, OS, Glob, Datetime, Seaborn, Matplotlib a widget Qgrid. Použití těchto knihoven a widgetu je popsáno v návrhu řešení.

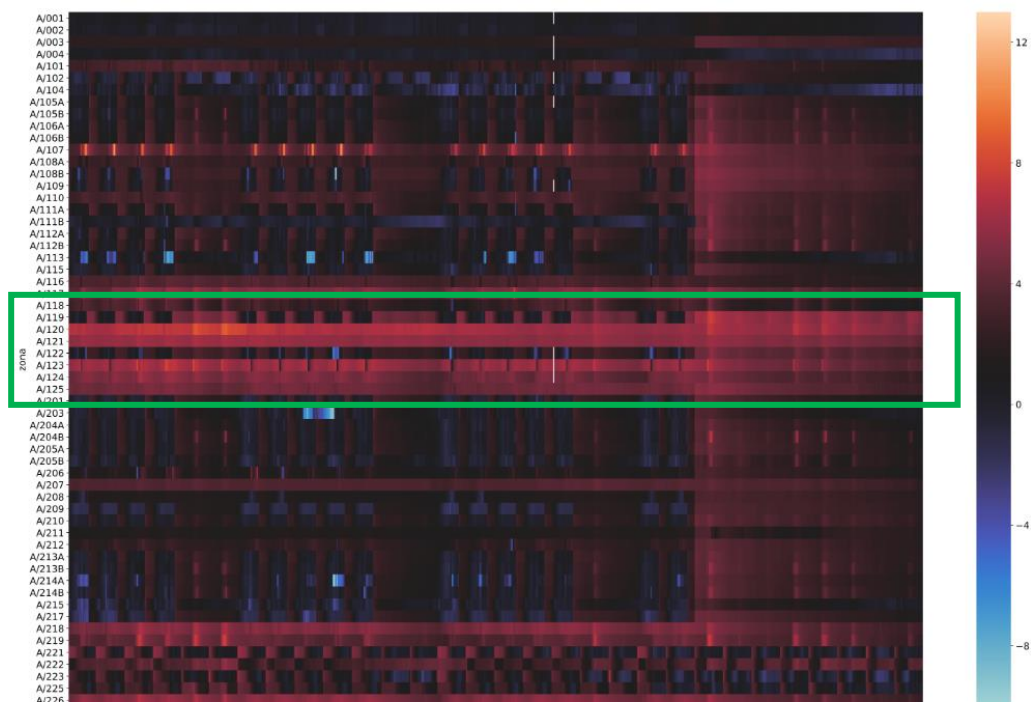
## 5. Experimentální výsledky

Interpretace výsledku analýzy v podobě vizualizace grafy, formulacemi zákonitostí a komentáři.

### 5.1 Přetápění a promrzání zón

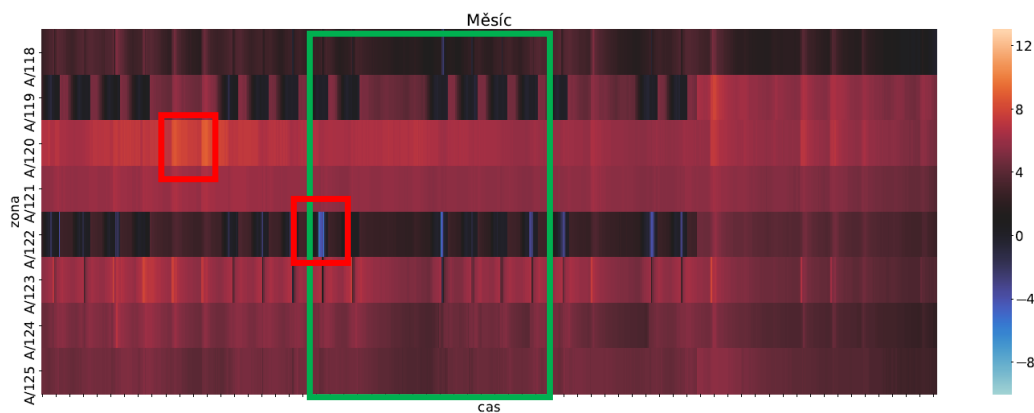
V prvním experimentu jsem prostřednictvím teplotní mapy graficky vizualizoval rozdíl skutečné a požadované teploty jednotlivých zón a na vzniklé teplotní mapě hledal, zda nedochází k přetápění nebo promrzání jednotlivých zón. Samotnou grafickou vizualizaci jsem rozdělil na časově omezené části, měsíc, týden, den a 6 hodin. Na první teplotní mapě (obrázek č. 36) vizualizuji všechny zóny zónového systému za měsíc prosinec roku 2015. Na teplotní mapě je černou barvou znázorněn ideální stav skutečné a požadované teploty nebo minimální odchylka. Červená barva znázorňuje přetápění naopak modrá barva znázorňuje

promrzání dané zóny. Čím světlejší odstín barvy červené nebo modré, tím je větší rozdíl mezi skutečnou a požadovanou teplotou.



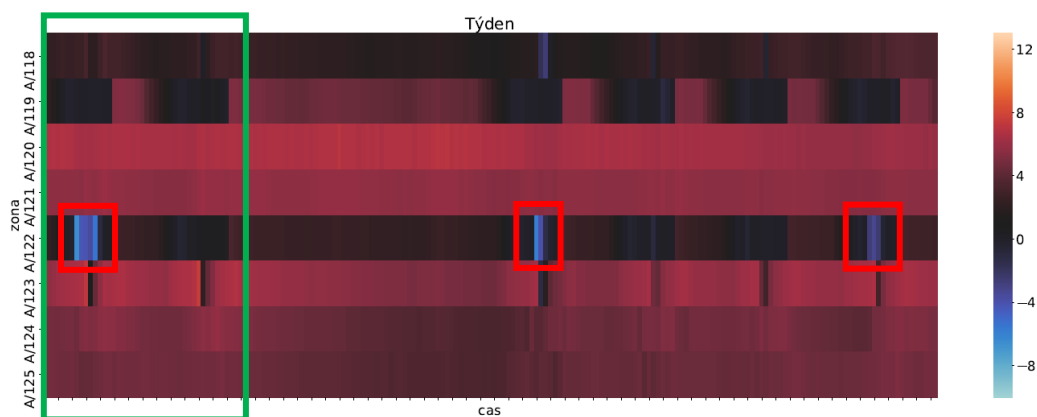
Obrázek č. 36 - Teplotní mapa rozdíl I.

Zóny A/001 a A/002 se chovají na rozdíl od zbylých zón konstantně a nevykazují výrazné odchýlení od požadované teploty. Pro bližší vizualizaci jsem zvolil několik zón, které vykazovaly z teplotní mapy nejvýraznější odchylky od požadované teploty a vizualizoval separátně (obrázek č. 37). Na teplotní mapě je vidět několik extrémních odchylek jak přetápění, tak promrzání ozn. červeně.



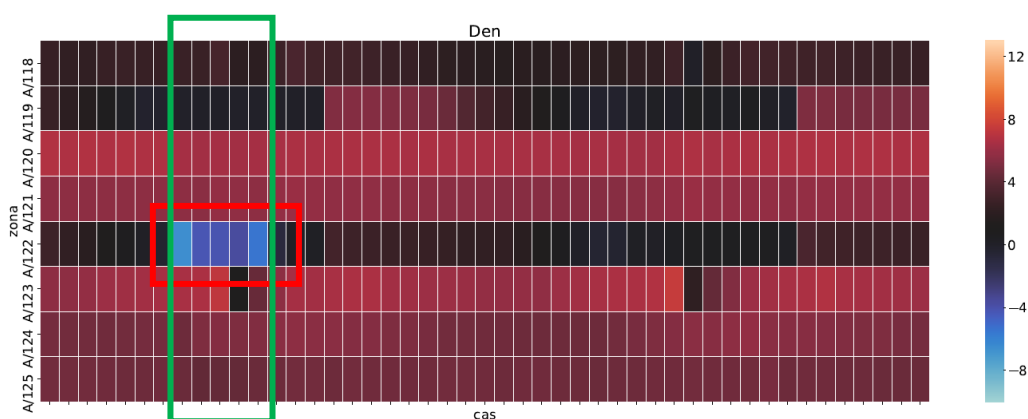
Obrázek č. 37 - Teplotní mapa rozdíl II.

Na dalším obrázku č. 38 vizualizují zeleně označenou část, která se zaměřuje na extrémní odchylky promrzání zóny v časovém omezení týden.



Obrázek č. 38 - Teplotní mapa rozdíl III.

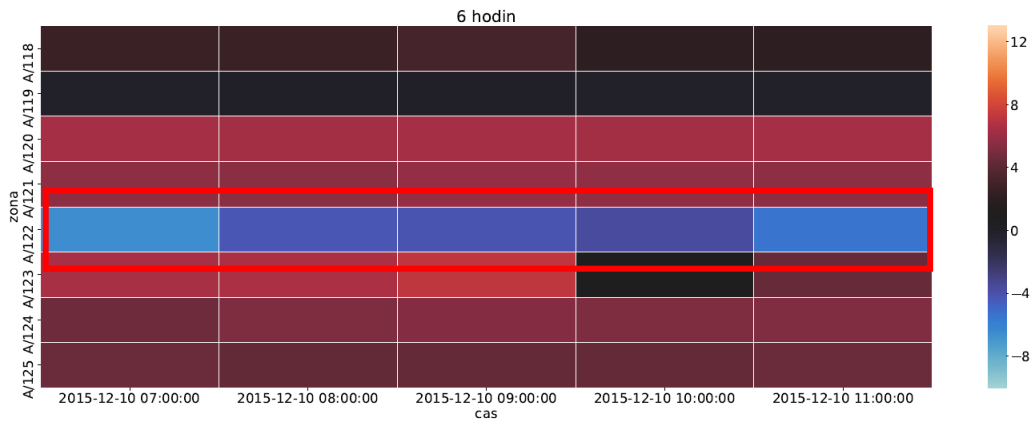
Na týdenním časovém omezení můžeme vidět, že zóna A/122 se chová poměrně konstantně, ale v několika případech dojde k extrémní odchylce promrzání. Na dalším obrázku č. 39 vizualizují zeleně označenou část v denním časovém omezení.



Obrázek č. 39 - Teplotní mapa rozdíl IV.

Na denním časovém omezení je vidět, že extrémní odchylka promrzání, ke které v dané zóně došlo, trvala několik hodin a rozdíl skutečné od požadované teploty byl v rozmezí  $-4$  °C až  $-8$  °C. Na poslední vizualizaci (obrázek č. 40) s přesností na hodiny určím kdy došlo k extrémní odchylce a můžu hledat příčinu, která jí způsobila.

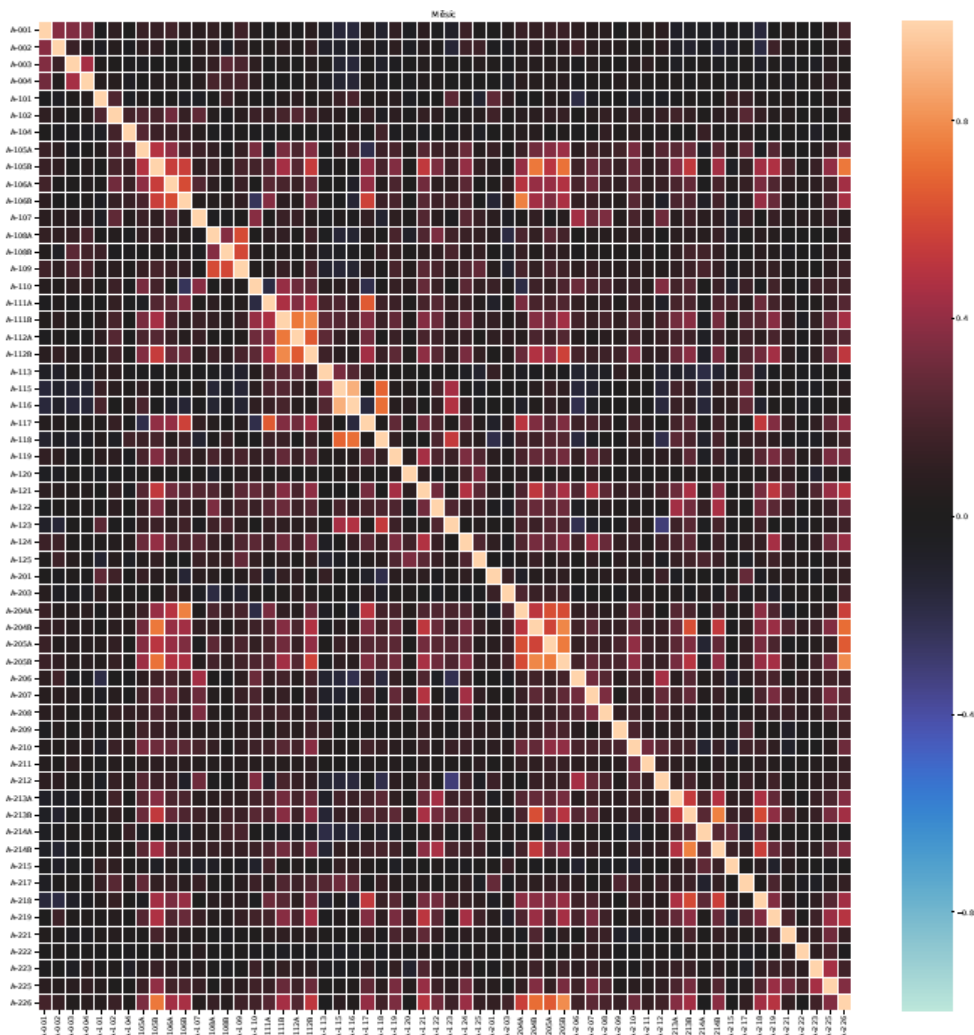




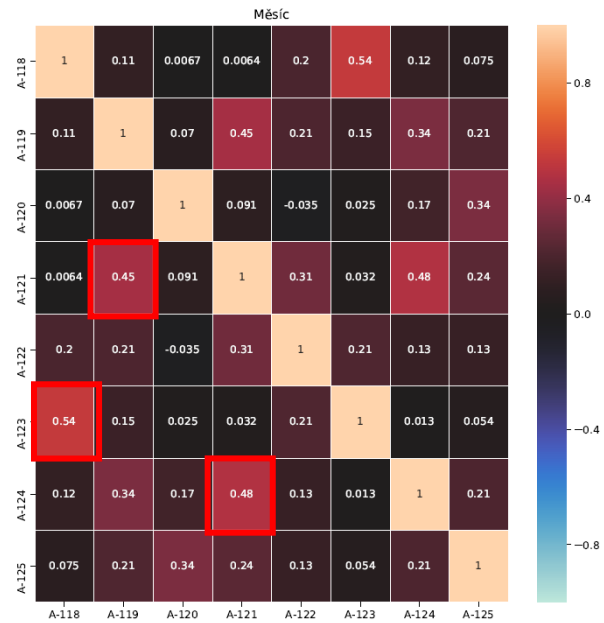
Obrázek č. 40 - Teplotní mapa rozdíl V.

## 5.2 Korelace jednotlivých zón

V druhém experimentu jsem prostřednictvím teplotních map graficky vizualizoval korelaci jednotlivých zón a na vzniklé teplotní mapě hledal zóny, které vzájemně korelují. Stejně jako v předchozím experimentu jsem grafickou vizualizaci rozdělil na časově omezené části, měsíc, týden, den a 6 hodin. Na první teplotní mapě (obrázek č. 41) vizualizuji všechny zóny zónového systému za měsíc prosinec roku 2015. Na teplotní mapě je znázorněna černou barvou nulová korelace nebo minimální korelace mezi dvěma zónami. Červená barva znázorňuje pozitivní korelaci naopak modrá barva znázorňuje negativní korelaci mezi dvěma zónami. Čím světlejší odstín barvy červené nebo modré, tím je větší síla korelace mezi dvěma zónami.

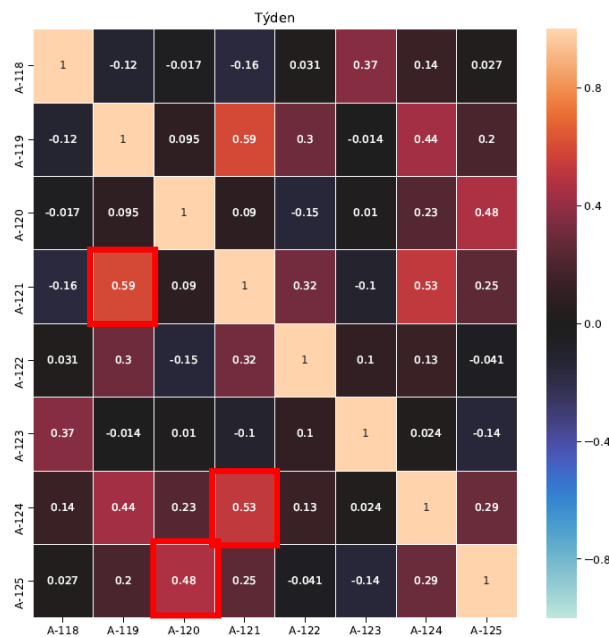


Obrázek č. 41 - Teplotní mapa korelace I.

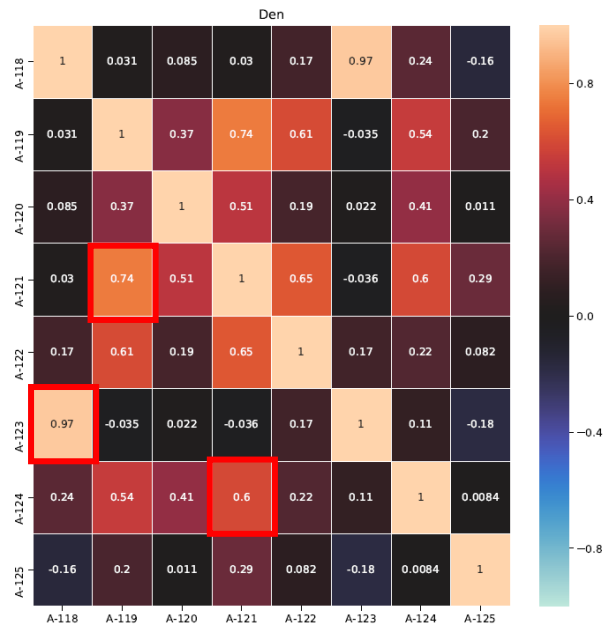


Obrázek č. 42 - Teplotní mapa korelace II.

Na teplotní mapách (obrázek č. 42 a 43) vizualizují korelaci jednotlivých zón v časovém omezení měsíc a týden nejvyšších hodnot korelace, a to pozitivní korelace dosahují zóny označené červenými značkami. Negativní korelaci v tomto časovém omezení dochází v minimálních hodnotách.

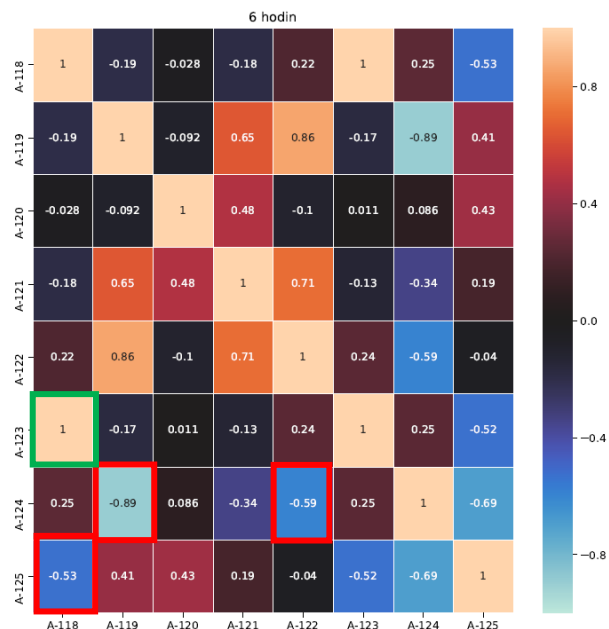


Obrázek č. 43 - Teplotní mapa korelace III.



Obrázek č. 44 - Teplotní mapa korelace IV.

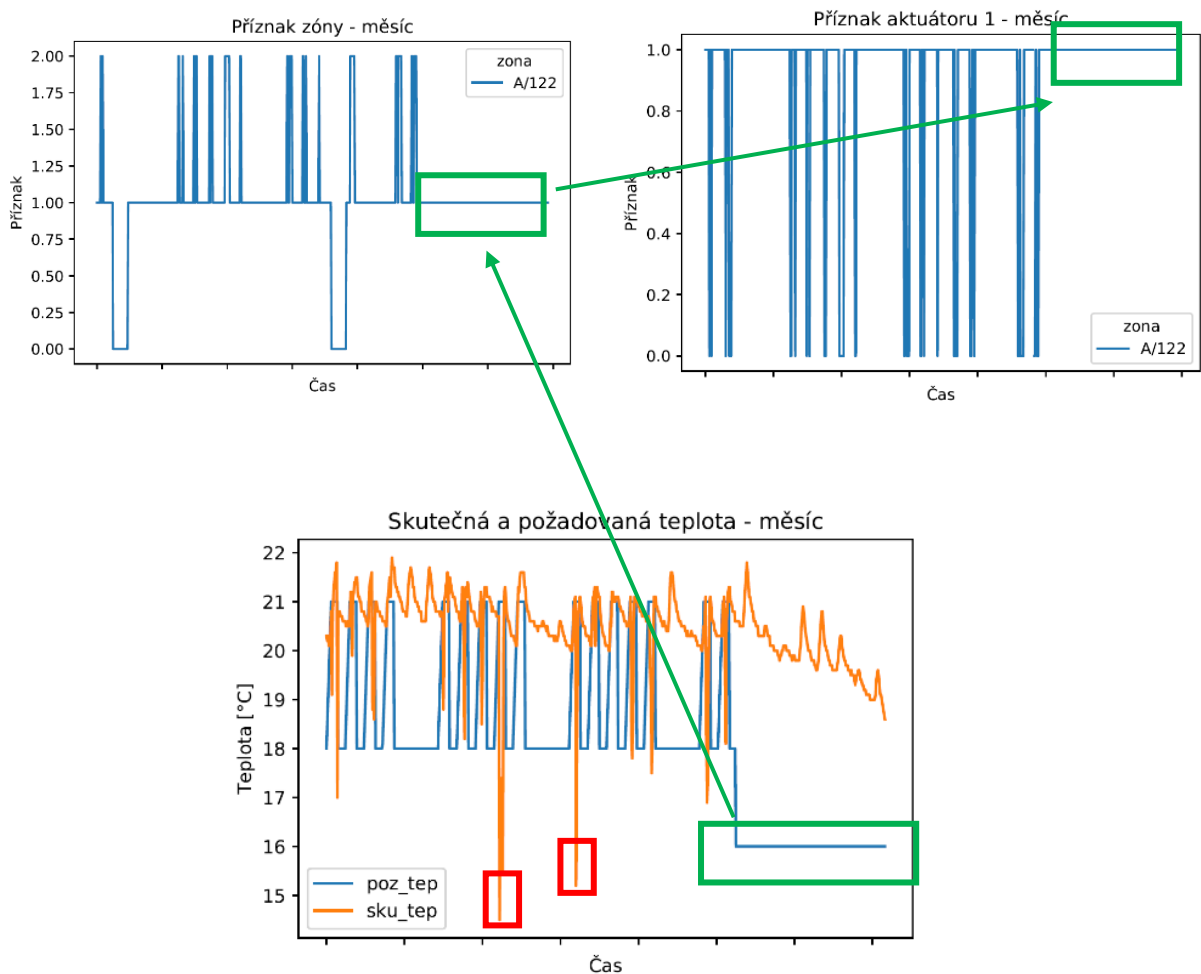
Na teplotní mapách (obrázek č. 44 a 45) vizualizují korelaci jednotlivých zón v časovém omezení den a 6 hodin. V obou případech dochází k pozitivní korelaci. Negativní korelaci s vysokými hodnotami dochází v časovém omezení 6 hodin, jednou z možných příčin může být rozdílná regulace zón. V jednom případě nastane i dokonalá pozitivní korelace, označena zelenou značkou.



Obrázek č. 45 - Teplotní mapa korelace V.

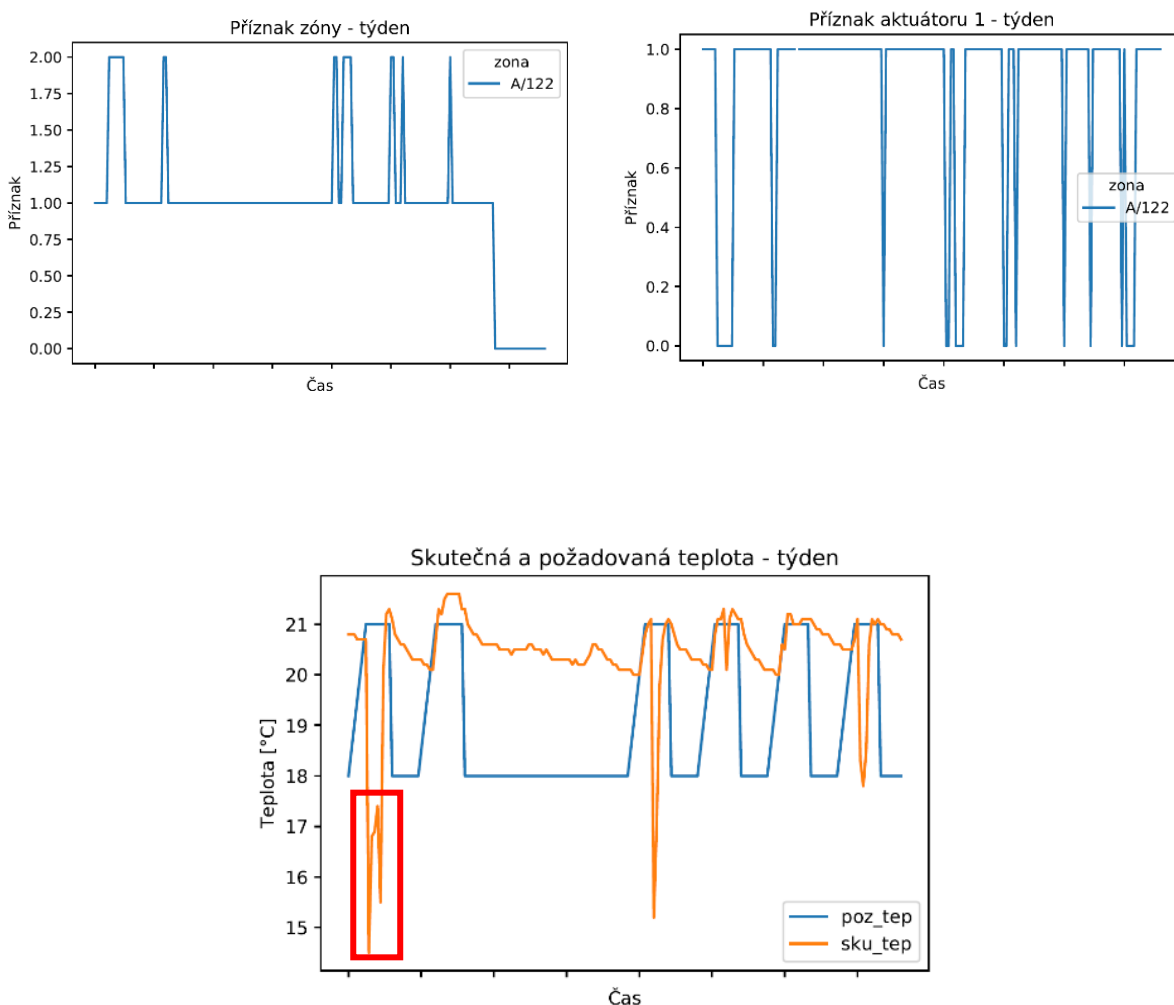
### 5.3 Chování zóny

Ve třetím experimentu jsem prostřednictvím spojnicového grafu vizualizoval průběh chování zóny A/122, u které došlo k extrémní odchylce, která byla odhalena prvním experimentem na teplotní mapě. V časových období jsem vizualizoval vždy tři typy grafů. Vizualizací příznaku zóny jsem sledoval, zda je daná zóna regulována nebo ne a pokud ano, o jakou regulaci se jedná zvyšování či snižování teploty. Vizualizací příznaku aktuátorů jsem sledoval, zda správně reaguje na příznak zóny stavem aktivním nebo neaktivním, protože to má vliv na topné těleso a na základě příznaku aktuátoru topné těleso topí či netopí. Vizualizací skutečné a požadované teploty jsem sledoval pohyb těchto teplot, zejména zda skutečná teplota výrazně klesá nebo stoupá nad požadovanou teplotu.



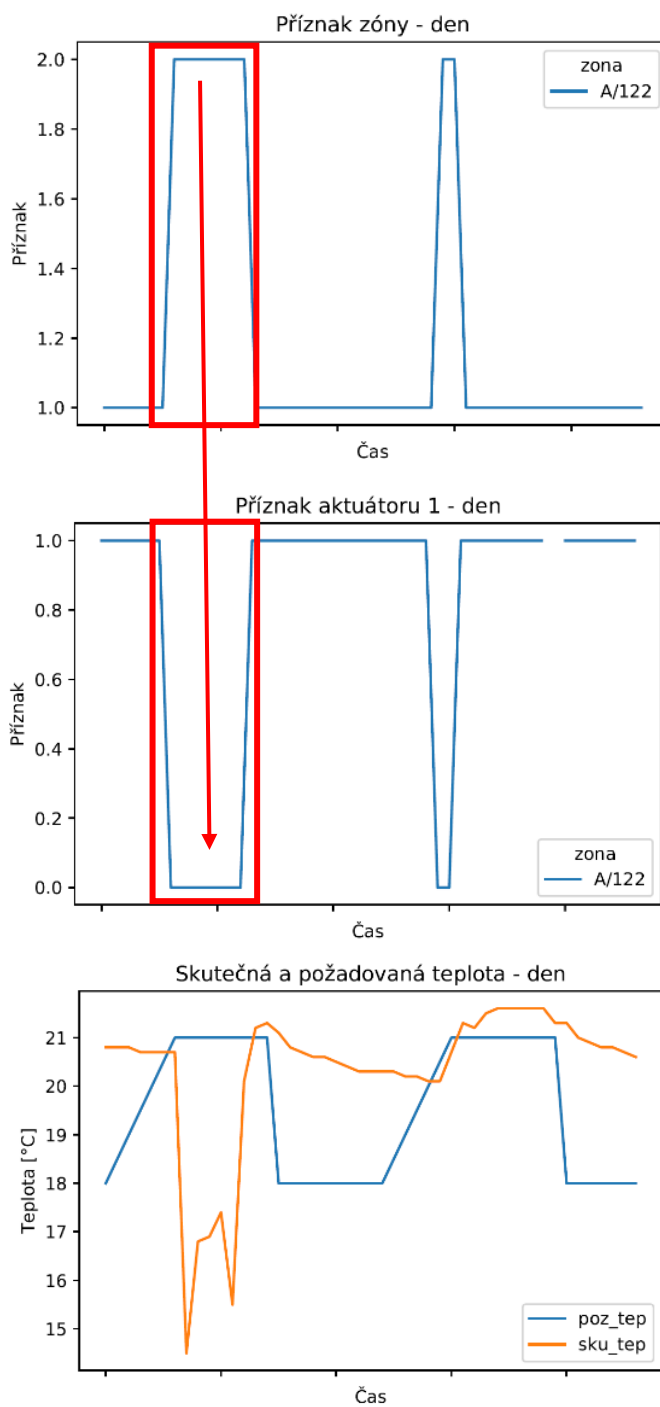
Obrázek č. 46 - Spojnicové grafy zóna I.

Z vizualizací v časovém rozmezí měsíc (obrázek č. 46), a to především z průběhů požadované teploty, lze vyčíst harmonogram využívání dané zóny. Při využívání zóny je nastavená požadovaná teplota na 21 °C jedná se o pracovní dny. Během víkendů a mimo dobu využívání je požadovaná teplota nastavena na 18 °C. Během svátku v prosinci daná zóna nebyla vůbec využívána a požadovaná teplota až do konce měsíce prosinec snížena na 16 °C na grafech zeleně označené, k poklesu skutečné teploty pod tuto hranici nedošlo a z toho důvodu nebylo nutné po celou dobu změnit stav aktuátorů do neaktivního stavu (0) tak, aby bylo zahájeno topení topného tělesa a zvyšování teploty. Přesto po celou dobu byla zóna ve stavu regulování, ale s příznakem snižování teploty (1). Červeně označené jsou extrémní odchylky, které jsem detekoval na teplotní mapě.



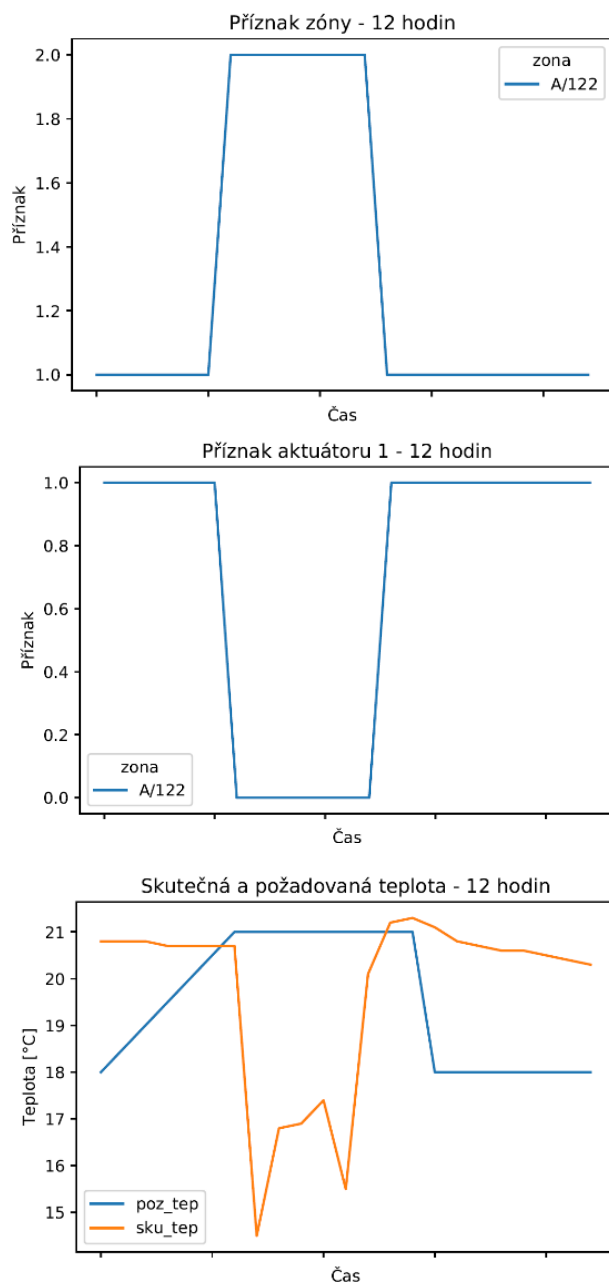
Obrázek č. 47 - Spojnicové grafy zóna II.

Z vizualizace v časovém rozmezí týden (obrázek č 47) zřetelně vidíme extrémní odchylky, ke kterým dochází v pracovní dny, kdy požadovaná teplota z důvodů využívání zóny je nastavena na hodnotu 21 °C. V dalším časovém rozmezí den a 12 hodin vizualizují červeně označenou odchylku.



Obrázek č. 48 - Spojnicové grafy zóna III.

Z vizualizace v časovém rozmezí den (obrázek č. 48 a 49) vidíme, že reakce zóny na stav skutečné a požadované teploty je správná. V zóně dochází k regulaci s příznakem zvyšování teploty (2) a na tuto skutečnost správně reaguje i aktuátor změnou příznaku na stav neaktivní (0), tím začne topné těleso topit a teplota v zóně by měla stoupat, přesto teplota klesá a výrazně. Skutečná teplota nad hranici požadované teploty vystoupá až koncem využívání zóny. Důvod, který tento stav způsobil není chybou individuální regulace zónového systému, ale způsoben jiným faktorem.

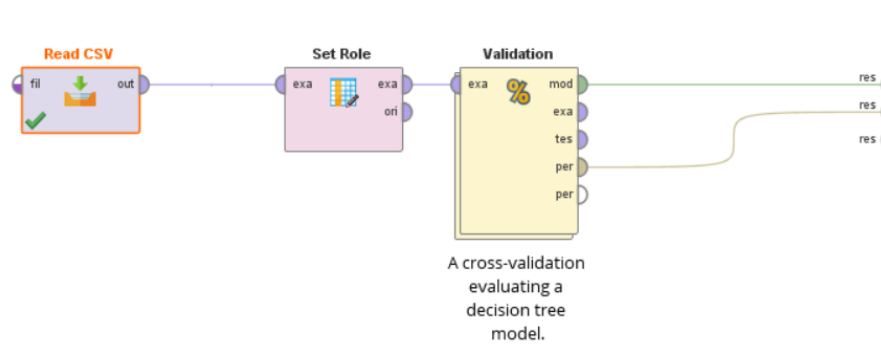


Obrázek č. 49 - Spojnicové grafy zóna IV.



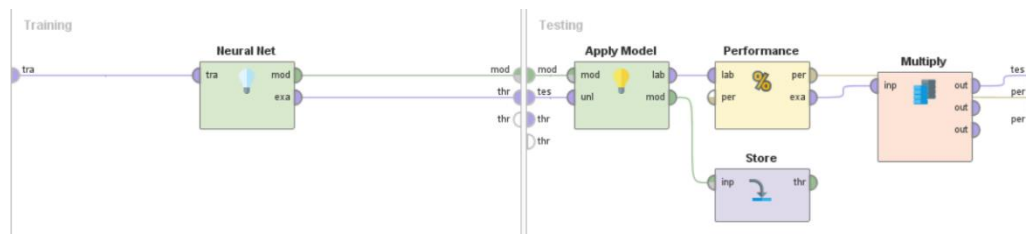
## 5.4 Predikce příznaku zóny

Ve čtvrtém experimentu jsem za pomoci neuronových sítí a chování zóny predikoval jeden ze tří příznaků zóny (tabulka č. 9). K experimentu jsem použil program RapidMiner. Data, která jsem použil k experimentu zahrnovala atributy, čas, skutečnou teplotu, požadovanou teplotu, příznak aktuátoru a zóny. V časovém rozmezí tří měsíců od začátku měsíce říjen do konce měsíce prosinec.



Obrázek č. 50 - Predikce příznaku zóny I.

Prvním krokem experimentu je načtení dat prostřednictvím modulu „Read CSV“ do programu RapidMiner a nastavení formátu. Důležitým a nezbytným krokem je nastavení tzv. „label“ atributu, který označuje predikovaný atribut to umožní modul „Set Role“. Posledním krokem v počátečním procesu (obrázek č. 50) je modul „Validation“, který vstupní data rozdělí na podmnožiny, kdy jedna podmnožina dat slouží jako testovací množina a zbylé podmnožiny jako trénovací množiny. Klasifikátor vytvoří model na trénovací množině a testuje ho pomocí testovací množiny. K opakování procesu křížové validace dochází několikrát s obměnou podmnožiny tvořící trénovací a testovací množinu. Proces křížové validace je realizován několika moduly (obrázek č. 51).



Obrázek č. 51 - Predikce příznaku zóny II.

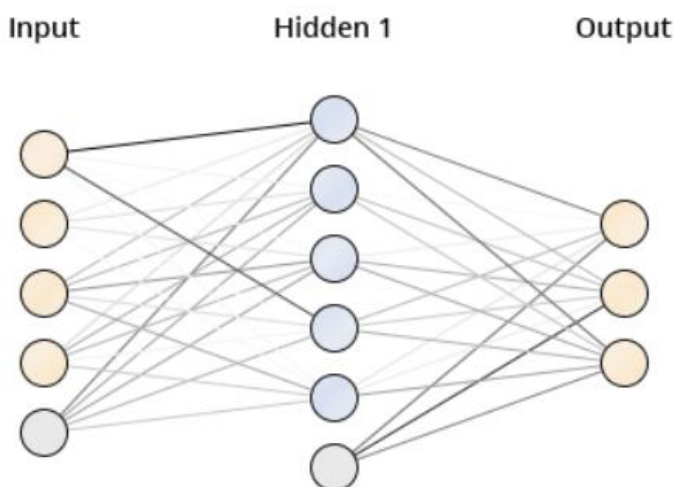
Klasifikátorem v tomto experimentu je neuronová síť, která realizuje proces učení bez učitele. V RapidMineru je realizována modulem „Neural network“ a umožňuje nastavení počtu skrytých vrstev, počtu neuronů ve skrytých vrstvách a počet trénovacích cyklů. Výsledkem učení neuronové sítě bez učitele je model, který je aplikován na data modulem „Apply model“ a tento model je uložen modulem „Store“ pro proces k učení neuronové sítě s učitelem. Po aplikování modelu na data dochází prostřednictvím modulu „Performance“ k výpočtu úspěšnosti učení neuronové sítě bez učitele neboli velikosti klasifikační chyby. Výsledek učení neuronové sítě s učitelem pokus první (obrázek č. 52).

accuracy: 94.51% +/- 1.32% (micro average: 94.51%)

	true NeReg	true RegSniTep	true RegZvyTep	class precision
pred. NeReg	902	4	4	99.12%
pred. RegSniTep	113	1652	11	93.02%
pred. RegZvyTep	21	6	183	87.14%
class recall	87.07%	99.40%	92.42%	

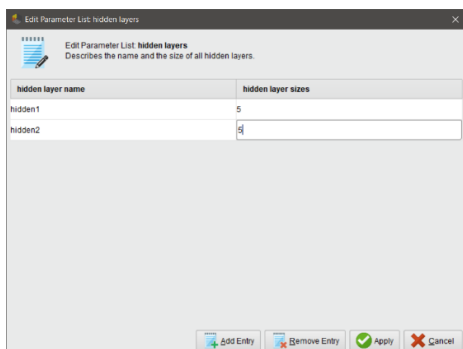
Obrázek č. 52 - Predikce příznaku zóny III.

Model učení neuronové sítě s učitelem pokus první (obrázek č. 53).

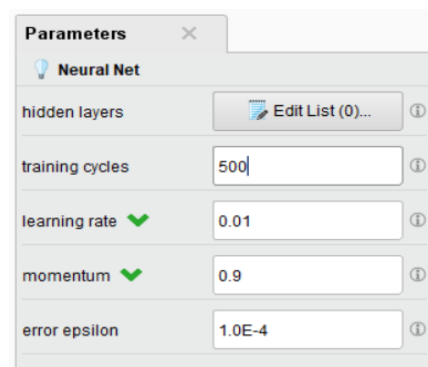


Obrázek č. 53 - Predikce příznaku zóny IV.

V experimentu jsem se pokusil dosáhnout lepších výsledků a pozměnil jsem počet skrytých vrstev z 1 na 2, počet neuronů v skrytých vrstvách ze 2 na 5 a počet trénovacích cyklů z 200 na 500 (obrázek č. 54 a 55).



Obrázek č. 54 - Predikce příznaku zóny V.



Obrázek č. 55 - Predikce příznaku zóny VI.

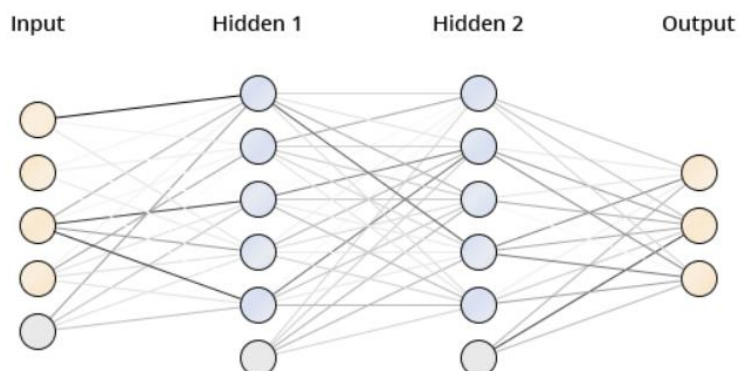
Výsledkem učení neuronové sítě s učitelem pokus druhý (obrázek č. 56).

accuracy: 94.48% +/- 1.34% (micro average: 94.48%)

	true NeReg	true RegSniTep	true RegZvyTep	class precision
pred. NeReg	902	4	4	99.12%
pred. RegSniTep	113	1650	10	93.06%
pred. RegZvyTep	21	8	184	86.38%
class recall	87.07%	99.28%	92.93%	

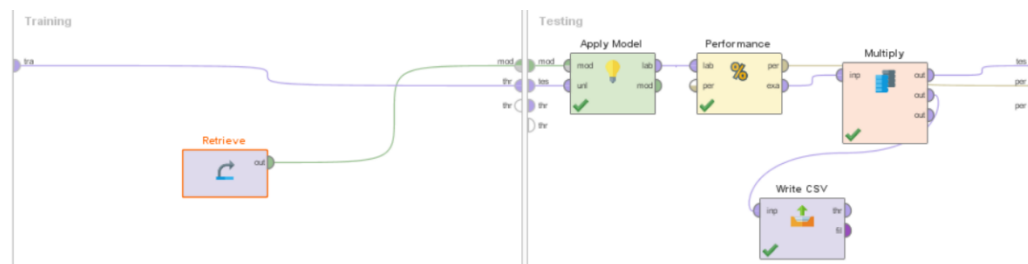
Obrázek č. 56 - Predikce příznaku zóny VII.

Model učení neuronové sítě s učitelem pokus druhý (obrázek č. 57).



Obrázek č. 57 - Predikce příznaku zóny VIII.

Z výsledků obou pokusů učení neuronové sítě s učitelem mohu konstatovat, že model s 2 skrytými vrstvami dosáhl od dvě desetiny horší výsledků. Pro další proces experimentu budu využívat prvního modelu neuronové sítě 4:5:3.



Obrázek č. 58 - Predikce příznaku zóny model I.

V poslední části experimentu (obrázek č. 58) načtu vytvoření model neuronové sítě z předchozího procesu učení neuronové sítě s učitelem modulem „Retrieve“. Načtený model aplikuju na data s cílem predikovat příznak zóny. Výsledkem po aplikování modelu (obrázek č. 59).

accuracy: 94.68% +/- 1.08% (micro average: 94.68%)

	true NeReg	true RegSníTep	true RegZvyTep	class precision
pred. NeReg	901	4	2	99.34%
pred. RegSníTep	113	1654	9	93.13%
pred. RegZvyTep	22	4	187	87.79%
class recall	86.97%	99.52%	94.44%	

Obrázek č. 59 - Predikce příznaku zóny model II.

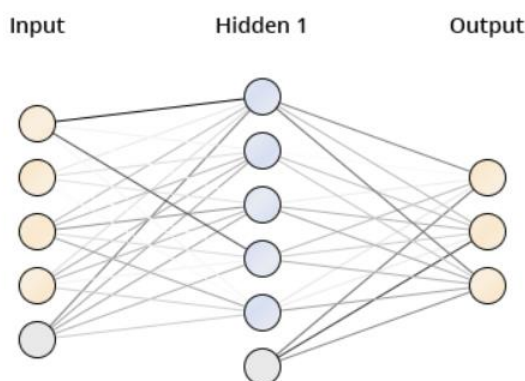
Pro lepší představu výsledku aplikovaného modelu na dat jsem provedl export dat do formátu csv (obrázek č. 60).

	A	B	C	D	E	F	G	H	I
1	cas	a1_pri	sku_tep	poz_tep	zona_pri	confidence(NeReg)	confidence(RegSníTep)	confidence(RegZvyTep)	prediction(zona_pri)
2	9/1/15 9:00 AM	1.0	29.3	21.0	NeReg	0.9554389488301888	0.04455692776106117	4.123408750036649E-6	NeReg
3	9/1/15 4:00 PM	0.0	29.0	21.0	NeReg	0.9660938801521924	0.03390277330505112	3.3465427563927426E-6	NeReg
4	9/2/15 3:00 AM	0.0	25.8	21.0	NeReg	0.9848670470860621	0.015129993475753611	2.959438184099601E-6	NeReg
5	9/2/15 8:00 AM	0.0	24.8	21.0	NeReg	0.9899426332383257	0.010054205681794479	3.1610798799127242E-6	NeReg
6	9/2/15 8:00 PM	0.0	25.0	21.0	NeReg	0.9891194706239994	0.010877443475685413	3.0859003151630707E-6	NeReg
7	9/3/15 4:00 PM	0.0	25.4	21.0	NeReg	0.987160252729367	0.012836755082765381	2.992187867610374E-6	NeReg
8	9/3/15 9:00 PM	0.0	24.6	21.0	NeReg	0.9906348379965558	0.009361909417501566	3.252585942664361E-6	NeReg
9	9/3/15 11:00 PM	0.0	24.5	21.0	NeReg	0.9909577152467866	0.009038976127456785	3.3086257565073492E-6	NeReg

Obrázek č. 60 - Predikce příznaku zóny model III.

Na základě výsledku experimentu je optimálním nastavením neuronové sítě pro predikci příznaku zóny 1 skrytá vrstva s 5 neurony a počet trénovacích cyklů nastaven na 200. Zvýšení

počtu skrytých vrstev, počtu neuronu v skrytých vrstvách a počet trénovacích cyklů nemělo význam, protože při mých pokusech nebylo dosaženo lepších výsledku než s původním nastavením neuronové sítě.



Obrázek č. 61 - Predikce příznaku zóny model IV.

Podoba výsledné neuronové sítě (obrázek č. 61) se skládá z 5 vstupních neuronů a z 6 neuronů ve skryté vrstvě, přičemž jak ve vstupní, tak ve skryté vrstvě je jeden neuron tzv. „threshold node“ neboli prahový uzel. Výstupní vrstva obsahuje 3 neurony. Při experimentu bylo dosaženo s odchylkou necelých 6% predikce příznaku zóny na základě čtyř rysů chování zóny.

Úspěšnost predikce příznaku zóny pomocí neuronové sítě s učitelem (obrázek č. 62).

accuracy: 94.51% +/- 1.32% (micro average: 94.51%)

	true NeReg	true RegSniTep	true RegZvyTep	class precision
pred. NeReg	902	4	4	99.12%
pred. RegSniTep	113	1652	11	93.02%
pred. RegZvyTep	21	6	183	87.14%
class recall	87.07%	99.40%	92.42%	

Obrázek č. 62 - Výsledek predikce příznaku I.

Úspěšnost predikce příznaku zóny pomocí aplikování vytvořeného modelu (obrázek č. 63).

accuracy: 94.68% +/- 1.08% (micro average: 94.68%)

	true NeReg	true RegSniTep	true RegZvyTep	class precision
pred. NeReg	901	4	2	99.34%
pred. RegSniTep	113	1654	9	93.13%
pred. RegZvyTep	22	4	187	87.79%
class recall	86.97%	99.52%	94.44%	

Obrázek č. 63 - Výsledek predikce příznaku II.

Závěrem z dosažených výsledků konstatuji, že experiment predikce příznaku zóny z rysů chování zóny byl úspěšný. Učení neuronové sítě s učitelem dosáhlo výsledku predikce 94,51% a aplikování vytvořeného modelu na data dosáhlo lepšího výsledku predikce 94,68%.

## 6. Závěr

Cílem této bakalářské práce byla analýza dat poskytnutých společnostmi zabývající se automatizací topných soustav inteligentních budov a vhodnými metodami modelovat chování topné soustavy a detekovat anomálie. Cíl práce, respektive všechny dílčí cíle byly splněny.

V úvodní části bakalářské práce byly definovány základní pojmy, detailně charakterizovány poskytnutá data a systém z něhož vznikly. Charakterizování dat jsem provedl od obecného popisu po charakterizování jednotlivých atributů a hodnot, kterých nabývají. U zónového systému jsem se zaměřil na jednotlivé části systému a princip, který jsem popsal slovně a doplnil schématem. Tím byl splnění první dílčí cíl.

V návrhu řešení bakalářské práce popisuji, jaký byl postup řešení, respektive co a jak bylo nutné udělat. Návrh řešení tvoří několik fází: export dat z databáze, selekce dat, předzpracování dat, transformace dat, dolování znalostí z dat a vizualizace dat. Počáteční fází byl export dat z databáze, protože poskytnutá data se nacházela na serveru v databázi InfluxDB. Po exportu dat a prvotních vizualizací jsem zjistil, že data nejsou v dostatečném formátu pro další zpracování a bylo nutné data předzpracovat. Předzpracování dat nebylo vůbec jednoduché, protože data se nacházela v jednom souboru, což způsobovalo nepřehlednost, a ne zcela dostačující vizualizaci, z toho důvodu jsem zařadil před fází předzpracování dat fází selekce dat. Selekcí dat jsem provedl rozdělení celku dat na jednotlivé zóny, to mi umožnilo zpracovávat každou zónu zvlášť a zaměřit se na korektnost záznamu dané zóny. Ve fází předzpracování dat jsem provedl změnu formátu času, selekci atributů, kontrolu počtu měření za každý den, doplnění chybějících měření, odstranění nadbytečných měření, sjednocení času měření, validaci měření, přejmenování atributů a odstranění nepotřebných atributů. Fáze předzpracování dat je základní zpracování dat, ale pro potřeby některých analytických metod nedostatečná. Fází transformací dat jsem doplnil základní zpracování o potřeby použitých analytických a vizualizačních metod. Transformace dat zahrnovala doplnění atributů, selekci hodnot atributů a selekci atributů. Ve fází dolování znalostí popisují použité analytické metody neuronové sítě a korelaci. Poslední fází je vizualizace, ve které popisují použité vizualizační metody teplotní mapy a spojnicové grafy. Každá tato fáze je podrobně popsána a je uvedeno, proč bylo zvoleno toto řešení. Tím byl splněn druhý dílčí cíl.

V implementaci řešení bakalářské práce popisuji, jakým způsobem jsou jednotlivé fáze návrhu řešení implementovány. Tím byl splněn třetí dílčí cíl.

V poslední části bakalářské práce prezentuji experimentální výsledky dosažené provedením analýz dat. Celkem jsem provedl čtyři experimenty. V prvním experimentu, Přetápění a promrzání zón, jsem analyzoval rozdíl skutečné a požadované teploty prostřednictvím vizualizace teplotní mapou. Na vytvořených teplotních mapách sledoval, zda nedochází k přetápění či promrzání zóny. Ve druhém experimentu, Korelace zón, jsem analyzoval korelaci jednotlivých zón prostřednictvím vizualizace teplotní mapou. Na vytvořených teplotních mapách sledoval, zda mezi sebou některé zóny korelují. Ve třetím experimentu, Chování zón, jsem analyzoval chování zóny vizualizací tří typů spojnicových grafů: skutečné a požadované teploty, příznaků zóny a příznaků aktuátoru. Na vytvořených spojnicových grafech sledoval, jak reagují příznaky zóny a aktuátoru na skutečnost, že skutečná teplota neodpovídá požadované teplotě. V posledním experimentu, Predikce příznaku zóny, jsem použitím neuronové sítě na základě atributů času, skutečné teploty, požadované teploty a příznaku aktuátoru predikoval příznak zóny. Tím byl splněn čtvrtý a poslední dílčí cíl.

Závěrem bych společnosti doporučil implementaci teplotních map do systému a analyzovat data z kotelny, které obsahují venkovní teplotu. Korelace venkovní teploty a jednotlivých zón by mohla poskytnou velice užitečná data.



## 7. Literární přehled

### 7.1 Knižní publikace

- [1] RUD, Olivia Parr. *Data Mining: praktický průvodce dolováním dat pro efektivní prodej, cílený marketing a podporu zákazníků (CRM)*. Praha: Computer Press, 2001, 329 s. Databáze. ISBN 80-722-6577-6.
- [2] BAŠTA, Jiří. *Regulace vytápění*. Praha: Vydavatelství ČVUT, 2002, 99 s. ISBN 80-010-2582-9.
- [3] SUMMERFIELD, Mark. *Python 3: výukový kurz*. Brno: Computer Press, 2010, 584 s. ISBN 978-80-251-2737-7.
- [4] PILGRIM, Mark. *Ponořme se do Python(u) 3: Dive into Python 3*. Praha: CZ.NIC, c2010, 434 s. CZ.NIC. ISBN 978-80-904248-2-1.
- [5] CHEN, Daniel Y. *Pandas for everyone: Python data analysis*. Boston: Addison-Wesley, [2018], 224 s. Addison-Wesley data and analytics series. ISBN 978-013-4546-933.
- [6] FANTA, Jiří. *Neuronové sítě ve společenských vědách*. Praha: Karolinum, 2000, 160 s. ISBN 80-246-0175-3.
- [7] TUČKOVÁ, Jana. *Vybrané aplikace umělých neuronových sítí při zpracování signálů*. Praha: České vysoké učení technické v Praze, 2009, 224 s. ISBN 978-80-01-04229-8.
- [8] MELOUN, Milan a Jiří MILITKÝ. *Statistická analýza experimentálních dat*. Vyd. 2., upr. a rozš. Praha: Academia, 2004, 960 s. ISBN 80-200-1254-0.

### 7.2 Manuály

- [9] *DOT smartDIRC manuál*. DOT CONTROLS a.s., 2010. 82 s.

### 7.3 Webové publikace

- [10] MATZ, Ph.D., Ing. Václav. Zónové regulační systémy a jejich využití při úsporném efektivním vytápění. TZBinfo [online]. 2010 [cit. 2018-04-17]. Dostupné z: <https://vytapeni.tzb-info.cz/mereni-a-regulace/6203-zonove-regulacni-systemy-a-jejich-vyuziti-pri-uspornem-efektivnim-vytapeni>

- [11] MATZ, Ph.D., Ing. Václav. Regulace teploty v budovách teoretická základna. TZBinfo [online]. 2010 [cit. 2018-04-17]. Dostupné z: <https://www.tzb-info.cz/4015-regulace-teploty-v-budovach-teoreticka-zakladna>
- [12] GSM. Wikipedie [online]. 2018 [cit. 2018-06-20]. Dostupné z: [https://cs.wikipedia.org/wiki/Global\\_System\\_for\\_Mobile\\_Communications](https://cs.wikipedia.org/wiki/Global_System_for_Mobile_Communications)
- [13] Grafana - The open platform for analytics and monitoring. Grafana - The open platform for analytics and monitoring [online]. Copyright © Grafana Labs [cit. 20.12.2018]. Dostupné z: <https://grafana.com/>
- [14] InfluxData (InfluxDB) | Time Series Database Monitoring & Analytics. InfluxData (InfluxDB) | Time Series Database Monitoring & Analytics [online]. Copyright © 2019 InfluxData, Inc. [cit. 20.12.2018]. Dostupné z: <https://www.influxdata.com/>
- [15] Home - Anaconda. Home - Anaconda [online]. Copyright © Copyright 2019 Anaconda, Inc. All Rights Reserved. [cit. 15.01.2019]. Dostupné z: <https://www.anaconda.com/>
- [16] Project Jupyter | Home. Project Jupyter | Home [online]. Copyright © 2019 Project Jupyter [cit. 15.01.2019]. Dostupné z: <https://jupyter.org/>
- [17] Lightning Fast Data Science Platform for Teams | RapidMiner©. Lightning Fast Data Science Platform for Teams | RapidMiner©[online]. Copyright © [cit. 15.01.2019]. Dostupné z: <https://rapidminer.com/>
- [18] GitHub - quantopian/qgrid: An interactive grid for sorting, filtering, and editing DataFrames in Jupyter notebooks. The world's leading software development platform · GitHub [online]. Copyright © 2019 [cit. 20.01.2019]. Dostupné z: <https://github.com/quantopian/qgrid>
- [19] Teplotní mapy pro web: porovnání nástrojů a k čemu slouží. Strategická reklamní agentura s digitálními kořeny: RobertNemec.com[online]. Copyright © Full service reklamní agentura RobertNemec.com, s. r. o. a Robert Němec 2001 [cit. 10.01.2019]. Dostupné z: <https://robertnemec.com/heatmapy-teplotni-mapy-porovnani/>
- [20] Korelační a regresní analýza – Wikisofia. [online]. Copyright © 2013 ISSN [cit. 25.05.2018]. Dostupné z: [https://wikisofia.cz/wiki/Korela%C4%8Dn%C3%AD\\_a\\_regresn%C3%AD\\_anal%C3%BDza](https://wikisofia.cz/wiki/Korela%C4%8Dn%C3%AD_a_regresn%C3%AD_anal%C3%BDza)

- [21] Co je Korelace | Peníze.cz. Peníze.cz - Půjčky, Kurzy měn, Akcie, Hypotéky, Bydlení, Daně [online]. Copyright © 2000 [cit. 17.04.2019]. Dostupné z: <https://www.penize.cz/slovník/korelace>
- [22] Correlation - Statistical Techniques, Rating Scales, Correlation Coefficients, and More - Creative Research Systems. Survey Software | Creative Research Systems [online]. Copyright © 2016 Creative Research Systems, All Rights Reserved [cit. 12.05.2018]. Dostupné z: <https://www.surveysystem.com/correlation.htm>
- [23] Positive Correlation. Investopedia - Sharper Insight. Smarter Investing. [online], [cit. 10.01.2019]. Dostupné z: <https://www.investopedia.com/terms/p/positive-correlation.asp>
- [24] Negative Correlation. Investopedia - Sharper Insight. Smarter Investing. [online], [cit. 10.01.2019]. Dostupné z: <https://www.investopedia.com/terms/n/negative-correlation.asp>
- [25] Artificial Neural Networks (ANN). Investopedia - Sharper Insight. Smarter Investing. [online], [cit. 10.01.2019]. Dostupné z: <https://www.investopedia.com/terms/a/artificial-neural-networks-ann.asp>

## 8. Seznam obrázků

Obrázek č. 1 - Schéma zónového systému .....	14
Obrázek č. 2 - Schéma návrhu řešení .....	18
Obrázek č. 3 - Prvotní náhled na data individuální regulace.....	19
Obrázek č. 4 - Prvotní náhled na data z kotelny .....	20
Obrázek č. 5 - Kód načtení dat .....	22
Obrázek č. 6 - Kód rozdělení dat I.....	23
Obrázek č. 7 - Kód rozdělení dat II. ....	23
Obrázek č. 8 - Kód rozdělení dat III. ....	23
Obrázek č. 9 - Kód uložení dat .....	24
Obrázek č. 10 - Kód načtení a uložení dat I. ....	24
Obrázek č. 11 - Kód načtení a uložení dat II. ....	25
Obrázek č. 12 - Kód formát času .....	25
Obrázek č. 13 - Kód selekce atributů .....	25
Obrázek č. 14 - Kód počet měření.....	26
Obrázek č. 15 - Kód čas měření I. ....	27
Obrázek č. 16 - Kód čas měření II.....	27
Obrázek č. 17 - Kód validace měření .....	27
Obrázek č. 18 - Kód přejmenování atributů .....	27
Obrázek č. 19 - Kód časové rozmezí I.....	28
Obrázek č. 20 – Kód časové rozmezí II.....	28
Obrázek č. 21 - Kód funkce časový rozsah .....	29
Obrázek č. 22 - Kód odstranění atributů.....	29
Obrázek č. 23 - Kód doplnění atributů I. ....	30
Obrázek č. 24 - Kód doplnění atributů II.....	30
Obrázek č. 25 – Kód zmapování hodnot atributů.....	30
Obrázek č. 26 - Spojení do celku I. ....	31
Obrázek č. 27 - Spojení do celku II. ....	31
Obrázek č. 28 - Spojení dle atributu I.....	32
Obrázek č. 29 - Spojení dle atributu II. ....	32
Obrázek č. 30 - Dopředná neuronová síť.....	34
Obrázek č. 31 - Učení s učitelem.....	35
Obrázek č. 32 - Učení bez učitele.....	36
Obrázek č. 33 – Implementace teplotní mapa rozdíl .....	36

Obrázek č. 34 – Implementace teplotní mapa korelace .....	37
Obrázek č. 35 - Implementace spojnicový graf .....	37
Obrázek č. 36 - Teplotní mapa rozdíl I.....	39
Obrázek č. 37 - Teplotní mapa rozdíl II. ....	39
Obrázek č. 38 - Teplotní mapa rozdíl III. ....	40
Obrázek č. 39 - Teplotní mapa rozdíl IV.....	40
Obrázek č. 40 - Teplotní mapa rozdíl V. ....	41
Obrázek č. 41 - Teplotní mapa korelace I. ....	42
Obrázek č. 42 - Teplotní mapa korelace II. ....	43
Obrázek č. 43 - Teplotní mapa korelace III.....	43
Obrázek č. 44 - Teplotní mapa korelace IV.....	44
Obrázek č. 45 - Teplotní mapa korelace V.....	44
Obrázek č. 46 - Spojnicové grafy zóna I. ....	45
Obrázek č. 47 - Spojnicové grafy zóna II. ....	46
Obrázek č. 48 - Spojnicové grafy zóna III.....	47
Obrázek č. 49 - Spojnicové grafy zóna IV. ....	48
Obrázek č. 50 - Predikce příznaku zóny I. ....	49
Obrázek č. 51 - Predikce příznaku zóny II. ....	49
Obrázek č. 52 - Predikce příznaku zóny III.....	50
Obrázek č. 53 - Predikce příznaku zóny IV.....	50
Obrázek č. 54 - Predikce příznaku zóny V.....	51
Obrázek č. 55 - Predikce příznaku zóny VI.....	51
Obrázek č. 56 - Predikce příznaku zóny VII. ....	51
Obrázek č. 57 - Predikce příznaku zóny VIII.....	51
Obrázek č. 58 - Predikce příznaku zóny model I. ....	52
Obrázek č. 59 - Predikce příznaku zóny model II. ....	52
Obrázek č. 60 - Predikce příznaku zóny model III.....	52
Obrázek č. 61 - Predikce příznaku zóny model IV.....	53
Obrázek č. 62 - Výsledek predikce příznaku I. ....	53
Obrázek č. 63 - Výsledek predikce příznaku II. ....	54

## 9. Seznam tabulek

Tabulka č. 1 - Popis atributů dat z individuální regulace .....	15
Tabulka č. 2 - Popis příznaků zóny .....	16
Tabulka č. 3 - Popis příznaků aktuátorů .....	16
Tabulka č. 4 - Popis atributů dat z kotelny .....	17
Tabulka č. 5 - Popis adres datových bodů .....	17
Tabulka č. 6 - Selekcce atributů .....	26
Tabulka č. 7 - Přejmenování atributů .....	28
Tabulka č. 8 – Příznaky aktuátorů .....	30
Tabulka č. 9 - Příznaky zóny .....	31

## 10. Seznam příloh

- [1] CD s textem bakalářské práce, programy k analýze dat v Jupyter Notebook a procesy v RapidMineru