



Report / SenseNet



Nadace
Vodafone
Česká republika

SenseNet Platforma Umělá inteligence

Sunday, July 02, 2017

Lichnovsky Adam

Version 1.0



Historie dokumentu

| Verze | Datum | Popis změn | Zodpovědná osoba za změny |
|-------|--------|-------------|-----------------------------|
| 1.0 | 7/2/17 | První verze | Adam_Lichnovsky@hotmail.com |

Obsah

1 Informace k projektu SenseNet, část Umělá inteligence

- 1.1 Kontakty
- 1.2 O projektu SenseNet
- 1.3 SenseNet Platforma

2 SenseNet platforma a umělá inteligence

- 2.1 Srovnání výstupů z modelu ALADIN a SenseNet UI
 - 2.1.1 ALADIN
 - 2.1.2 SenseNet UI

3 SenseNet UI Platforma

- 3.1 Popis problému
- 3.2 Časové řady
- 3.3 Výběr algoritmů
 - 3.3.1 Tradiční analytické algoritmy
 - 3.3.2 Analytické algoritmy strojového učení
- 3.4 4 důležité faktory při výběru algoritmů strojového učení
 - 3.4.1 SenseNet UI algoritmy
 - 3.4.2 Neuronová síť
- 3.5 SenseNet UI Neuronové sítě
- 3.6 Vývoj modelu neuronové sítě
- 3.7 Infrastruktura SenseNet UI Platformy
 - 3.7.1 Vizualizace designu komponent a modulů
 - 3.7.2 Požadavky na SenseNet UI infrastrukturu
 - 3.7.3 SenseNet UI Instalované prostředí

4 Stav odvedené práce

- 4.1 Presentace úvodu k Neuronovým sítím
- 4.2 Identifikace algoritmů
- 4.3 Vývoj modelu neuronové sítě
 - 4.3.1 malyTeplomer

5 Plán

1 Informace k projektu SenseNet, část Umělá inteligence

1.1 Kontakty

| Jméno a příjmení | Role | Telefonní číslo | Kontaktní email |
|------------------|------------|-----------------|-----------------------------|
| Adam Lichnovský | AI Inženýr | +420728051380 | Adam_lichnovsky@hotmail.com |

1.2 O projektu SenseNet

Stav životního prostředí na Ostravsku má přímý dopad na kvalitu života, zdraví obyvatel. Vnímáme potřebu integrace a získávání aktuálních, transparentně dostupných dat. Chceme vytvořit otevřenou platformu využitelnou státní správou, neziskovým sektorem, komerční sférou i koncovými uživateli. Využití pokročilé umělé inteligence umožní zpracování a analýzu dat v reálném čase a na základě těchto informací můžou zainteresované strany změnit svoje chování s pozitivním dopadem na životní prostředí, kvalitu života a snížení nákladů na energie. Řešení může být dále využitelné i do jiných regionů.

1.3 SenseNet Platforma

SenseNet platforma je informační systém, který bude podporovat a pomáhat s řešením různých enviromentálních problémů v ostravském regionu. Jedním z benefitů platformy je využití algoritmů umělé inteligence, která umožní širokou škálu analýzy a zpracování dat. Hlavním zdrojem dat pro SenseNet platformu jsou převážně veřejně dostupné zdroje měřených veličin ovzduší, jako jsou teplota, vlhkost, tlak vzduchu a také koncentrace škodlivin ve vzduchu. Pomocí algoritmů umělé inteligence se provedou analýzy a předpovědi nad měřenými meteorologickými daty, které dále využije SenseNet platforma ve svých ostatních komponentách. Primární využití výstupních dat umělé inteligence budou zpracovávat aplikace pro vzdálené řízení ovládacích prvků dostupných IoT zařízení a aplikace pro detekci kvality ovzduší.

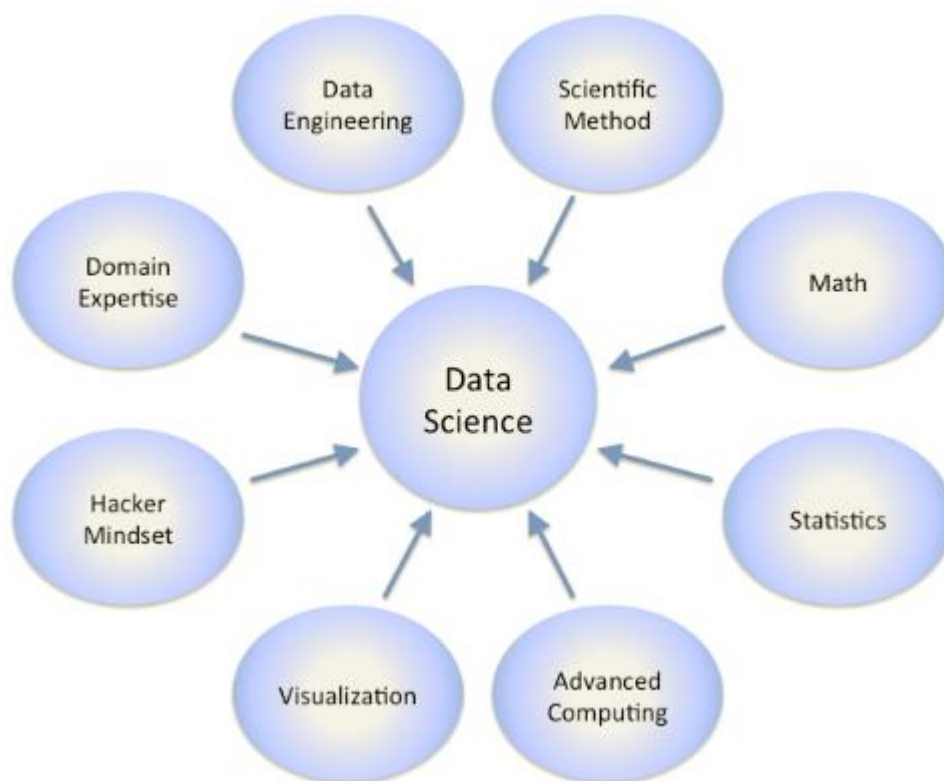
2 SenseNet platforma a umělá inteligence

SenseNet umělá inteligence bude představovat schopnost platformy předpovídat a proaktivně reagovat na neustále se měnící ovzduší na Ostravsku.

Pro tyto účely se SenseNet Umělá inteligence bude skládat z metod a algoritmů z oblastí strojového učení (machine learning) a vědy hromadných údajů (Data Science).

Strojové učení je podoblastí umělé inteligence, zabývající se algoritmy a technikami, které umožňují počítačovému systému 'učit se'. Učením v daném kontextu rozumíme takovou změnu vnitřního stavu systému, která zefektivní schopnost přizpůsobení se změnám okolního prostředí.

Strojové učení prochází v posledních několika letech vzrušujícím obdobím, které lze implementovat velice rychle do stávajících informačních systémů. Využití algoritmů strojového učení sníží výpočetní požadavky na předpovědi v porovnání na výpočetní požadavky deterministických matematických modelů (např. ALADIN).



2.1 Srovnání výstupů z modelu ALADIN a SenseNet UI

Uvedme si příklad, jak nám mohou algoritmy strojového učení pomoci při analýze dat a jejich předpovědi ve srovnání s předpověďmi modelu ALADIN.

2.1.1 ALADIN

- Centralizovaná platforma: nejdříve se ve středisku Météo-France v Toulouse vypočte méně podrobný globální model ARPÉGE. Jeho výsledky se pak přenesou do jednotlivých členských zemí a zde se v modelu ALADIN zpřesňují pro konkrétní území.
- Vysoké nároky na výpočetní výkon.
- Hlavní výsledky jsou k dispozici **4× denně** – v 0, 6, 12 a 18 h. – a předpovídají počasí vždy na následujících **54 hodin**.
- Výstupem (výsledkem) modelu jsou předpovědi řady fyzikálních parametrů atmosféry, z nichž se kreslí např. nejznámější 4 typy předpovědních map:
 - a) teploty ve 2 m nad zemí,
 - b) srážky,
 - c) směr a rychlost větru,
 - d) oblačnost.

2.1.2 SenseNet UI

- Centralizovaná platforma: UI zpracovává dostupná měřená meteorologická data. Přesto otevřená implementace SenseNet Platformy umožní širší distribuci a další vývoj předpovědních modelů.
- Hlavní výsledky jsou k dispozici **24× denně** – předpovídají počasí vždy na následujících **24 hodin**.
- Výstupem (výsledkem) modelů SenseNet UI jsou předpovědi řady fyzikálních parametrů atmosféry:
 - a) teploty ve 2 m nad zemí,
 - b) směr a rychlost větru,
 - c) smogové znečištění (koncentrace SO_x, NO_x, PMO částice).

3 SenseNet UI Platforma

3.1 Popis problému

V úvodu jsme si již nastínili problém, který chceme pomocí SenseNet Platformy a jejích komponent řešit. Pojdme si problém ovzduší a předpovědi o jeho stavu projít hlouběji. Řekněme si, že máme dostupná data z různých zdrojů, které zajišťuje komponenta ‚crawlers‘ SenseNet Platformy. Data v platformě jsou stahována a ukládána do databází. Dále jsou meteorologická data předzpracována do jednotlivých souborů dat, které pak následně zpracovává umělá inteligence platformy.

Hlavním úkolem SenseNet AI bude zpracování dat naměřených časových řad. Údaje o časových řadách budou představovány naměřenou teplotou, vlhkostí vzduchu a koncentrací znečištění ve vzduchu. Znečištění bude představováno koncentrací prachových částic, oxidů síry, oxidů uhlíku a nitro-oxidů ve vzduchu. Data o znečištění zpracovávané umělou inteligencí jsou časové řady. Problematiku časových řad si přiblížíme v další kapitole.

3.2 Časové řady

Analýza časových řad je stále jedním z obtížných problémů v oblasti vědy o datech a je zajímavou oblastí výzkumu. Existuje mnoho příkladů časových řad kolem nás. Předpovědi počasí, ceny energie, prognózy prodeje nebo předpovědi ceny akcií společnosti Vodafone pomocí deterministických modelů jsou velmi náročné na výpočetní výkon, který vypočítává proměnné rovnice. Stochastická povaha meteorologických událostí v naměřených časových řadách činí řešení problému velmi obtížným.

Časové řady často obsahují skryté časové závislosti, které způsobují stav, kdy dva jinak identické časové okamžiky, do kterých patří různé třídy naměřených dat, předpovídají odlišné chování. Tato charakteristika obecně zvyšuje rozdíly v jejich analýze. Stávající techniky často závisí na ručně vytvořených vlastnostech, které byly drahé a vyžadovaly odborné znalosti v oboru. S nástupem strojového učení byly vyvinuty nové modely využívající podobnosti funkcí buněk neuronů pro analýzu časových řad a prognózy. Výsledky ukazují, že strojové učení má k tomu hodně přispět.

3.3 Výběr algoritmů

3.3.1 Tradiční analytické algoritmy

Tradiční analýza časových řad zahrnuje rozložení dat do jejich složek, jako je trendová složka, sezónní složka a šum. Techniky jako ARIMA (p, d, q) model, klouzavý průměr, automatické regrese budou také použity k analýze časových řad.

3.3.2 Analytické algoritmy strojového učení

Odpověď na otázku „Jaký algoritmus strojového učení máme použít?“ je vždy „To záleží“. Záleží na velikosti, kvalitě a povaze dat. Záleží na tom, co chcete s odpovědí dělat. Záleží na tom, jak byla matematika algoritmu přeložena do pokynů pro počítač, který používáte. Záleží na tom, kolik času máte. Dokonce i zkušení vědci v oblasti dat neumějí zjistit, který algoritmus bude nejlépe fungovat předtím, než je vyzkouší.

3.4 4 důležité faktory při výběru algoritmů strojového učení

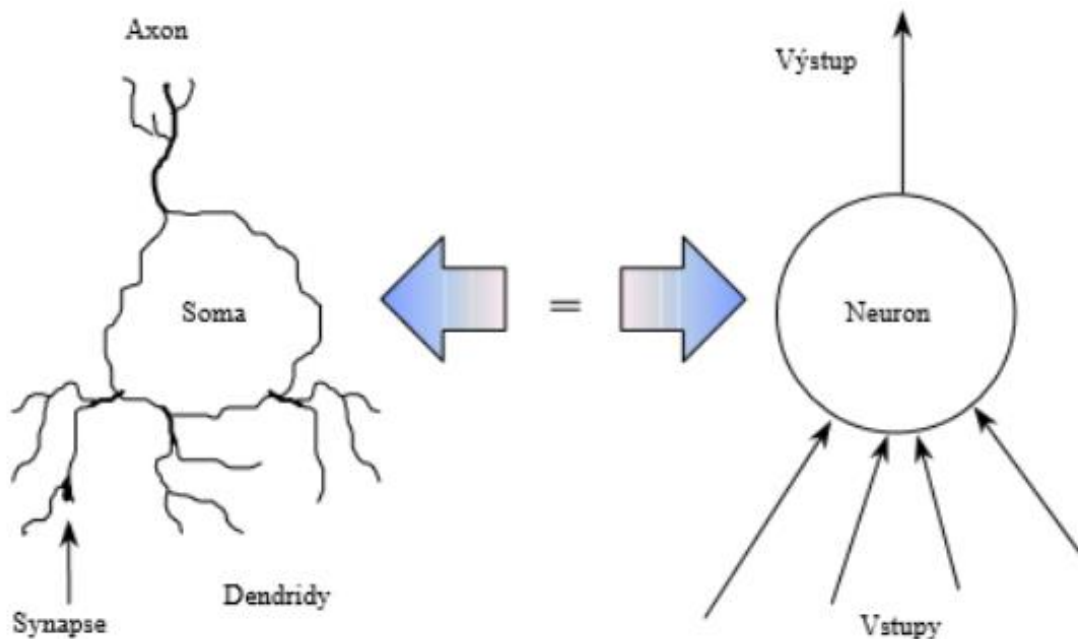
1. Výpočetní výkon (Moorův zákon, GPUs, ASICs, atd.)
2. Data (připravené ve strukturované formě, ne uložené jen tak někde - např. ČHMÚ)
3. Algoritmy (zkoumej a hledej: např. BackPropagation, FNN, CNN, LSTM)
4. Infrastruktura (Linux, TCP/IP, Git, ROS, PR2, AWS, AMT, TensorFlow, Python atd.)

3.4.1 SenseNet UI algoritmy

Po počáteční analýze byly vybrány vhodné algoritmy pro strojové učení, které budou implementovány a používány SenseNet UI platformě. Jedním z hlavních modelů jsou modely neuronových sítí. Možnosti modelů neuronových sítí, další možnosti přizpůsobení funkcí neuronových sítí. Funkce neuronových sítí nám pomohou v blízké budoucnosti rozšířit možnosti regrese a předpovědí dat časových řad.

3.4.2 Neuronová síť

Neuronová síť je jeden z výpočetních modelů používaných v umělé inteligenci. Jejím vzorem je chování odpovídajících biologických struktur. Umělá neuronová síť je struktura určená pro distribuované paralelní zpracování dat.



Neuronová síť se skládá z umělých (nebo také formálních) neuronů, jejichž předobrazem je biologický neuron. Neurony jsou vzájemně propojeny a navzájem si předávají signály a transformují je pomocí určitých přenosových funkcí. Neuron má libovolný počet vstupů, ale pouze jeden výstup.

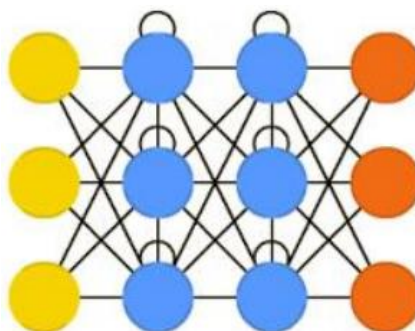
3.5 SenseNet UI Neuronové sítě

SenseNet UI Analýzy časových řad se budou provádět hlavně pomocí následujících algoritmů, které se ukázaly jako velmi účinné v analýze časových řad v posledních letech:

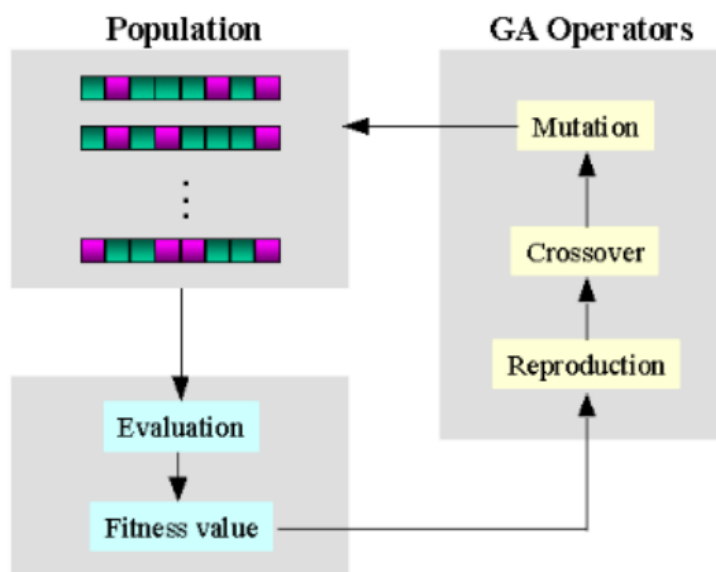
- Feed Forward Neuronových sítí,



- Rekurentní neuronové sítě, jako je Long-Short Term Memory,



- Genetické algoritmy, využité zvláště pro optimalizaci předpovědí.



3.6 Vývoj modelu neuronové sítě

Vývoj modelu neuronových sítí a jejich optimalizace v několika krocích:

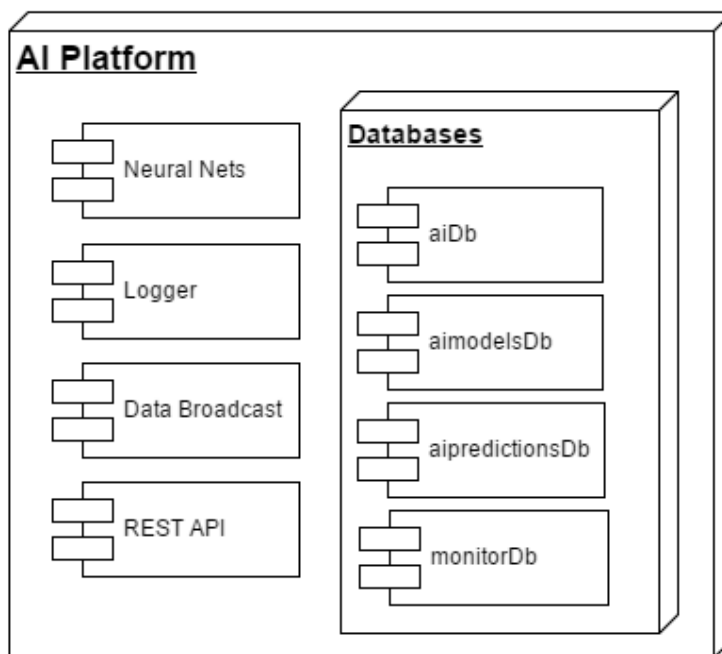
1. Získej data pro neuronovou síť.
2. Připrav strukturovaná data pro neuronovou síť.
3. Připrav model neuronové sítě.
4. Vytrénuj model neuronové sítě.
5. Vyhodnoť chybovost modelu neuronové sítě.
6. Model neuronové sítě předpovídá budoucí data.
7. Vyhodnoť výkonost modelu neuronové sítě.
8. Optimalizace kódu modelu neuronové sítě.
9. Model neuronové sítě předpovídá budoucí data s vysokou výpočetní efektivitou.

3.7 Infrastruktura SenseNet UI Platformy

UI moduly – **Neural Nets**, **Logger**, **Data Broadcast**

UI databáze - **aiDb**, **aimodelsDb**, **aipredictionDb**

3.7.1 Vizualizace designu komponent a modulů



3.7.2 Požadavky na SenseNet UI infrastrukturu

1. **Požadovaný hardware:** Ověřeno pro použití s více generacemi procesorů Intel a kompatibilních procesorů, mimo jiné včetně: procesoru Intel® Xeon®, rodiny procesorů Intel® Core™ a rodiny procesorů Intel® Atom™.
2. **Operační systém:** Použito stejné API pro vývoj aplikací na více operačních systémech: Windows, Linux a OS X.
3. **Vývojářské nástroje a prostředí:** Eclipse s PyDev (Windows, Linux a OS X).
4. **Programovací jazyky:** Primárně podporujeme vývoj v Pythonu a při druhotné pokročilé optimalizaci v jazycích C a C++.

3.7.3 SenseNet UI Instalované prostředí

1. Hardware: CPU virtuální AMD_64.
2. Operační systém: CentOS 7.
3. Běžové prostředí: Python 3.5.3 a související balíčky (NumPy, SciKit, TensorFlow, Keras, Theano, PsiberLogic, matplotlib, IPython Jupyter, Pandas, SymPy).

4 Stav odvedené práce

4.1 Prezentace úvodu k Neuronovým sítím

Ve dnech 18. a 20. dubna proběhla přednáška „Úvod do neuronových sítí“, která přispěla k širšímu pochopení tematiky mezi řešiteli a veřejností.

[Úvod do neuronových sítí](#)

[Úvod do neuronových sítí vol. 2](#)

[Prezentace z přednášky](#)

4.2 Identifikace algoritmů

Proces výběru algoritmů probíhá stále formou čtení odborných článků a zkoušení funkčnosti designovaného konceptu daného algoritmu. Pokud se v čase ukáže prostor pro implementaci a testování, pak mezi již vybranými algoritmy přibudou další.

4.3 Vývoj modelu neuronové sítě

Byl vyvinut model neuronové sítě `malýTemplomer.py` typu Feed Forward. Tato síť při správném nastavení parametrů velikosti sítě a velikosti datového setu pro učení poskytuje výsledky s nízkým využitím výpočetního výkonu.

4.3.1 malyTeplomer

Struktura dat: časová řada

Zdroj dat: „dronova koupelna“

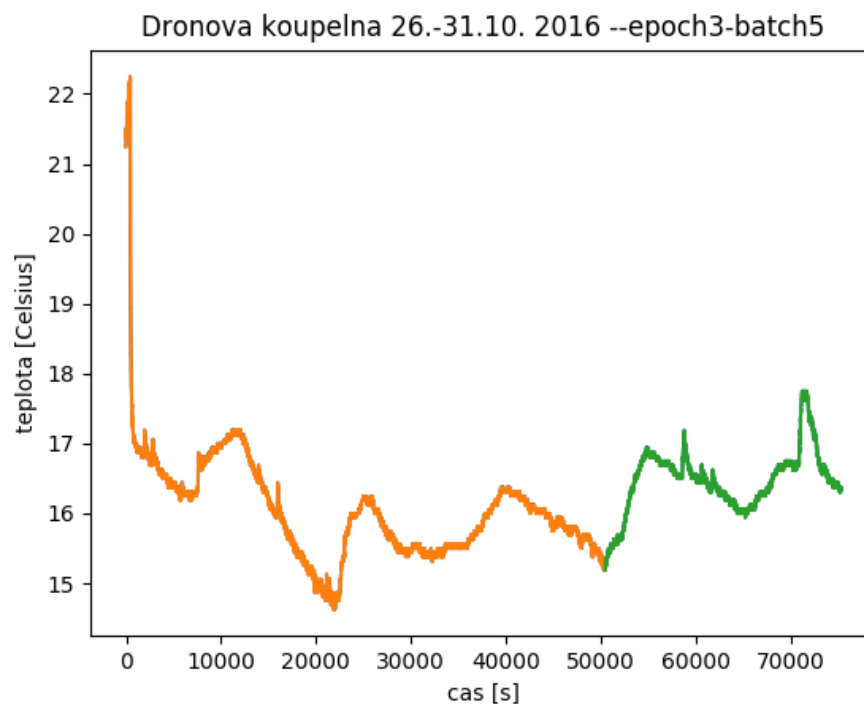
Samplování: 10 s

Model Neuronové sítě: Sekvenční (FeedForward)

Aktivační funkce: ReLU

Optimalizace: ADAM

Rozsah dat: Použitý datový set je mezi daty 2016-10-26 21:20:39 a 2016-10-31 23:59:59



Naměřený výkon:

Epocha 1/3: 8 s – loss: 4.7518

Epocha 2/3: 7 s – loss: 0.0013

Epocha 3/3: 8 s – loss: 9.5030e-04

Train Score: 0.00 MSE (0.03 RMSE)

Test Score: 0.00 MSE (0.03 RMSE)

Legenda:

- Oranžová barva znázorňuje učící sadu dat.
- Zelená barva znázorňuje predikované hodnoty teploty.

5 Plán

V následujících měsících se bude provádět následující práce:

- Modifikace FNN modelu neuronové sítě pro vícenásobná vstupní data,
- testování LSTM modelů neuronových sítí a optimalizace modelu,
- pokračování a zintenzivnění již započaté práce na API pro Umělou inteligenci,
- testování stávajících modelů neuronových sítí na posbíraných datech ‚crawlers‘,
- analýza výkonu modelů neuronových sítí v produkčním prostředí,
- automatizace „přípravy“ sbíraných dat pro modely neuronových sítí,
- integrace SenseNet UI komponenty s ostatními komponenty SenseNet Platformy.