1. Bevezető (2 oldal)
2. Irodalom feldolgozás, háttér információk (10%) 7oldal
3. A feladat megoldásához rendelkezésre álló technikák, technológiák, fejlesztőeszközök ismertetése, összehasonlítása, (előnyök, hátrányok ütköztetése, felhasználhatóság mérlegelése). (25%) 17oldal
4. Saját teljesítmény előállításához ténylegesen felhasznált eszközök részletes ismertetése; a kiválasztás szempontjai, telepítés, használatba vétel előfeltételei és lépései (ha azok nem magától értetődők). (20%)14oldal
5. Saját munkánk (alkalmazásfejlesztés, mérés, tervezés stb.) részletes leírása, az eredmények szemléletes ismertetése. (35%) 25oldal
6. Összegzés

Tapasztalatok, eredmények összegzése; a feladat elkészítése során felmerült nehézségek, elkövetett hibák; jövőbeni fejlesztési lehetőségek. (10%)7oldal

1. Bevezetés (1-2 oldal, a hallgató ismerteti a dolgozat témáját, feladatát és lehetséges módszereit)
2. Főtéma ismertetése (tématerület meghatározása és kifejtése, Fogalomkör meghatározása, vállalat bemutatása)
3. Probléma felvetése és kifejtése
4. Modellek, módszerek és megoldási elvek ismertetése
5. Számítások, rajzok, megoldások ismertetése és kifejtése a feladaton keresztül
6. Végleges megoldás meghatározása és kifejtése
7. Javaslatok további fejlesztésekre, esetleg más (jobb) megoldások ismertetése
8. Összegzés

összesen = 70oldal

Diplomatervezés 1:

1. Az osztályozási módszerek áttekintése
2. A problémakör információelemeinek elemzése
3. Reprezentációs módszerek, változatok vizsgálata
4. Adattisztítási lépések
5. Adatelemzési folyamat megtervezése

|  |  |
| --- | --- |
| Miskolci Egyetem  Gépészmérnöki és Informatikai Kar  Általános Informatikai Intézeti  *3515 Miskolc-Egyetemváros* |  |

**SZAKDOLGOZAT**

*Feladat címe:*

Mesterséges intelligencia alapú oktatást támogató alkalmazás tervezése

*Készítette:*

Szilvási Péter

*MSc szintű mérnökinformatika szakos*

*Alkalmazásfejlesztő szakirányos hallgató*

*Témavezető:*

Dr. Kovács László

*tanszékvezető, egyetemi tanár*

**Miskolc, 2020.**

# Bevezetés

* Általános bevezetés (miről szól a szakdolgozat?)
* Motiváció (miért kell támogatni az oktatást)
  + Kevés tanár
  + Még annál is kevesebb jó tanár
  + Sokféle tananyag -> nehéz megfelelő tanulási módszert kitalálni
  + Rengeteg tananyag (information overflow) -> sok köztük a fals(hamis) információk (misinformation): **ÖTLET**: Material filter (for misinformation)
  + Diákok különböző tanulási módszertanai vannak
  + Diákok nem azonos gyorsasággal tanulnak
  + Különféle tanulási típusok:
    - vizuális
    - auditív
    - verbális
    - mozgásos
    - logikai
    - intraperszonális (magányos)
    - interperszonális (társasági)
  + ...
  + Előnyök, hátrányok a típusok és módszerek között
    - vizuálishoz kivetítő -> eszköz hiány -> helyettesíteni mással
    - interperszonáls -> mai világba nehezebb sokkal szeparáltibbak vagyunk mind tanulás mind szociálisan -> **ÖTLET**: Student Group ***Clustering***/Classification: a társasági típusok között csoportokat alkotni, akik jól kijönnek + tanulnak egymással -> effektív tanulás + diák elégedettség + öröm
  + Milyen az én tanulási módszerem?
    - vizuális, mozgásos, logikai, inter/intra (függ a tananyagtól)
* Oktatás fontossága
  + Az oktatás az alapja mindenek, ennek ellenére nem fordítunk rá kellő figyelmet. Okai:
    - Nincs rövidtávú „haszna”. Hosszú távú (pl. matek, fizika felfedezések csak évtizedekkel később lehet alkalmazni).
    - Nem termel profitot, nem értékesít terméket.
    - Előítélet: könnyű szakmának gondoljuk, de a legnehezebb:  
      „Anyone can be a teacher or professor, but not everyone can influence you to strive for excellence and make a difference in the world around you.” Curtis White
* Átvezetés:
  + Már rengeteg MI applikáció van, viszont kevés foglalkozik az oktatással. Hogyan tudd az MI segítséget nyújtani az oktatásban?
  + Nehézségek: (Miért nehéz megállapítani a tanulási módszerünket?) nehéz megállapítanunk még a saját magunk legjobban illeszkedő tanulási módszert
    - Hiányzik az oktatásban tanulás módszer óra (Hogyan kell tanulni?)
    - Nincs időnk megismerni a dolgokat. Túl sok tananyag, túl sok számonkérés, (sok memórizáció).
    - Teljesítmény (jegy) orientált -> kevés kísérletezés. Mivel csak egy adott tananyag halmazból tanulunk -> 1szemszögből(perspektíva) alakítunk véleményt
  + Hogyan állapíthatjuk meg egy diák legjobban passzoló tanulási technikáját? Mesterséges intelligencia ezen a téren támogatást nyújthat nekünk.
  + Milyen lehet a neked legjobb tanulási módszer? Derítsük ki!

# Problémakör

* Adat szükségesége: bármilyen modell építéséhez adat kell (adat az új üzemanyag)
* Miért kellenek adatok egy MI modellhez?
  + Ember-gép kapcsolat -> embernek is kell gépnek is kell
  + Ember gondolkodás még máig nem értjük
  + Mai gépek működését is nagyon nehéz (sokak számára black box)
  + Matematika alapja. Koordináta rendszer, függvények, pontok
  + Adat = koordináta pont
  + Minél több koor. pont -> annál pontosabb függvény leírás -> pontosabb jóslat
* Machine learning pipeline (kép)
* Data pipeline kifejtése
* Többi pipeline kifejtése (model learning, model evaluation, model maintenance)
* Mik lehetnek adatok
  + személy attr., pénzügyi, weboldal használat, pixel
* Az alkalmazás esetében az adat
  + Általános iskola, Gimnázium, Egyetem
  + mind más és más tananyag
  + tanulási módszer: gyökerében hasonlók (kis eltérések)
  + ….
* …
* Adatforrás probléma: Honnan szerezhető adatforrás?
  + Kaggle, google dateset search, amazon, UCI, stb.
  + Akár készíthető is (kérdőívek, IOT rendszerek, szenzorok, stb.)
* Általános adatforrás leírása
  + null értékek
  + zajok
  + nem minden érték számosított
  + …
* Az alkalmazás elkészítése során használt adathalmaz ismertetése:
  + dimenzió (oszlopok, leírása)
  + mennyiség (hány darab sor)
  + …
  + Átvezetés: Hogyan tudunk összefoglaló képet kapni az adatunkról? A legkönnyebb módja a képi illusztráció. (rím átvezetés) ->

Többet mond egy kép  
mint száz bekezdés

# Reprezentációs módszerek

* Miért jó a reprezentálni/vizualizálni az adatot? Vizualizáció = megértés
* Vizualizációs kapacitás: Ember maximum 3d képes megérteni (azt is nehezen) inkább 2d a kézenfekvő. A gép bármennyi dimenziót képes feldolgozni
* 2D ábrázolás: Scatter, Grid, Plot, Heatmap, …
* python matpotlib
* 3D ábrázolás: Wire, Surface, Mesh, …
* Sok dimenzió ábrázolás problémája: Hogyan tudunk több dimenziós adatokat kisebb dimenzióba ábrázolni?
* …
* Átvezetés: Miért jó részletesen ismerni a tanulás során felhasznált adatokat?

# Adathalmazok ismertetése

* Fogalmak (adatspecifikus) tisztázása
  + input
  + output
  + dimenzió
  + feature
  + …
  + zaj (adatban)
  + adattisztítás
  + dimension reduction
  + feature selection
* Kevés dimenzióval -> nem jól korrelál (felszínes)
  + előny: gyors tanulás, kevés adat, nincs szükség v. kicsi adattisztításra, könnyű adatbővítés (kevesebb kérdés), …
  + hátrány: nem robosztus modell, …
* Sok dimenzió
  + előny: robosztus modell, …
  + hátrány: lassú tanulás, adattisztítás (zaj szűrés), nehéz adatbővítés, …
* Választani sok/kevés dimenzió között:
  + kevés dimenzió: jól megértett adat -> tudjuk mi a fontos a többi csak zaj
* …
* Fogalmak (statisztika specifikus) https://realpython.com/python-statistics/
  + std (standard deviation)
  + mean (weighted?), median, mode, avg,
  + variability (joint, corr)
  + variance
  + sűrrőség
  + eloszlás
* Amikkel dolgoztam adathalmaz ismertetése
  + kis dimenziós (8oszlopos)
  + nagy dimenzió (sok oszlopos…)
* Tapasztalat
  + statisztikák, (jó sok kép!!)
  + python kód (jupyter)
  + magyarázat

# Adattisztítási lépések

* Az adat statisztikai elemzése után -> adattisztítás
* Többféle tisztítás ismertetése
  + drop: kidobni sorkat/oszlopkat (ált. id oszlop kidobása)
  + replace null value -> 0 v. átlagértéket megadni stb.
  + …

# Modell ismertetése (tanulás, értékelés (evaluation) -> train/test)

[Hasonló struktúra mint az adathalmazok ismertetésénél -> következő félévben]

* Fogalmak
* …
* Ismertetni az emberi tanulást vs gép tanulást. Emberi tanulásból meríteni ihletet.
* …