

## 5. Stretnutie 31.10.2019

### Agenda

1. Dohodnúť sa na terminológii v rámci UI
2. Diskutovať pripravené návrhy diagramov

### Priebeh stretnutia

Spomenuli sme doteraz spravenú prácu v šprinte.

#### **Podobné autá zhluková analýza:**

- Na základe vloženého inzerátu vieme nájsť množinu podobných áut. Užívateľovi potom vieme zobrazíť inzeráty ktoré sú oveľa lacnejšie/drahšie a zobrazíť v akých parametroch sa líšia
- Zobrazíť škálu
- Bude treba zistiť podľa akých parametrov zhlukovať podobné autá. Ako jedna alternatíva sa zdá byť VIN číslo, ktoré obsahuje veľa kľúčových informácií o vozidle

#### **Predikcia ceny od dňa predaja**

Zobrazenie krivky, ktorá popisuje ako sa vyvíja cena podobných áut v čase.

Aj tu sme zistili, že je nutné vedieť dobre zhlukovať podobné autá

Zásadné je vedieť dobre identifikovať podobné autá - vedieť dobre zhlukovať. Preto musíme preskúmať dáta a nájsť atribúty, ktoré najviac korelujú s cenou auta. Tieto atribúty budú mať v zhlukovaní väčšiu váhu.

Krivka najazdených kilometrov ku roku výroby. Vieme povedať, či je auto zachovale/ normal/ veľmi ojazdené na svoj vek,  
Krivka vyvoja ceny auta po najazdených kilometroch resp. rokoch (bude tam nejaka korelacia).

Máme dataset z bazosu pre prácu.

V prípade podobnosti - zhluky, bolo by fajn mať možnosť ovládať veľkosť zhlukov.

Bavili sme sa o metódach identifikácie atribútov z textu. Plánujeme vyžiadať si prístup k NLP4SK od M. Bištáka a dohodnúť si s ním stretnutie, kde by sme sa ho chceli opýtať na jeho tipy pre náš typ úlohy.

Tiež sme sa bavili o možnom použití Word2Vec. Toto treba bližšie preskúmať.

Spravili sme backlog grooming.