MVD okruhy ke zkoušce – 2022

1. **Týden (Úvod)**
    a. Normalizace textu
    b. N-gramy
2. **Týden (Vizualizace dat)**
    a. Projít jednotlivé grafy – jaké informace z nich lze vyčíst
    b. Vizualizace hierarchií
    c. Redukce dimenze dat (PCA, T-SNE)
3. **Týden (Word2Vec)**
    a. Word2Vec, GloVe – jak funguje
    b. CBOW a Skip-gram princip trénování
    c. Softmax vs Negative Sampling
4. **Týden (Vyhledávání 1)**
    a. Jaké úlohy vyhledávač obsahuje
    b. Jak funguje indexování
    c. Vector Space Model – k čemu je potřeba
    d. TF-IDF, BM25
    e. Precision, Recall, F-Measure
5. **Týden (Vyhledávání 2)**
    a. Kosinova podobnost
    b. Jak vylepšit vyhledávání s Word2Vec
    c. Jaké jsou další možnosti vylepšení vyhledávání
    d. Siamská architektura neuronových sítí a její použití
6. **Týden (Hodnocení a vyhledávání na webu)**
    a. Z čeho se skládá vyhledávač
    b. K čemu je MapReduce
    c. PageRank, HITS
7. **Týden (Shlukování)**
    a. Co a k čemu to je, příklad aplikace
    b. Hierarchické shlukování
    c. K-Means
    d. Určení optimální hodnoty K u K-Means
8. **Týden (Shlukování 2)**
    a. DBSCAN
        i. Jak funguje
        ii. Core, border a outlier
        iii. Přímá a nepřímá dosažitelnost, propojenost
9. **Týden (Moderní jazykové modely)**
    a. Rekurentní neuronové sítě
    b. ELMO
    c. Transformer architektura
    d. BERT
10. **Týden (Aplikace jazykových modelů)**
    a. RoBERTa, ALBERT, ELECTRA, DistilBERT
    b. GLUE Benchmark – k čemu slouží
    c. Jak realizovat analýzu sentimentu nebo odpovídání na otázky s využitím BERT modelu

11. **Týden (Doporučovací systémy)**
    a. Content-based filtering
    b. Collaborative filtering
12. **Týden (Detekce anomálií)**
    a. Isolation forest
    b. Local Outlier Factor
13. **Týden (Vyhledávání vzorů)**
    a. Časté vzory, support, confidence
    b. Apriori algoritmus
    c. FPGrowth