Okruhy, ze kterých si vylosujete teoretickou otázku ke zkoušce `M8DM1 Data mining I`:

1. Data mining, KDD proces a vše kolem.
2. Dataminingové metodologie a příbuzné pojmy.
3. Relační databáze a jazyk SQL.
4. BI - datový sklad, OLAP, OLTP.
5. Integrace dat.
6. Čištění dat.
7. Transformace dat.
8. Redukce dat.
9. Vizualizace dat a informací.

Matematické (metodologické) otázky ke zkoušce `M8DM1 Data mining I`:

1. **Analýza hlavních komponent**. Popište matematický model analýzy hlavních komponent. Podrobně ukažte, jak jsou komponenty konstruovány. V čem spočívá redukce dimenze? Jak se interpretují její výsledky? Jak se v praxi aplikuje?

2. **Faktorová analýza a mnohorozměrné škálování**. Popište cíle a matematický model faktorové analýzy. Popište metody, jak se faktory hledají. Jak se v praxi aplikuje? Jak se interpretují její výsledky? K čemu slouží rotace? Popište matematický model metrického mnohorozměrného škálování. Jak se provádí zobrazení dat v prostoru nízké dimenze?

3. **Exploratorní analýza dat numerických dat**. Popište k čemu slouží exploratorní analýza dat. Detailně popište metody jednorozměrné a mnohorozměrné exploratorní analýzy pro numerické proměnné. Zaměřte se na číselné i grafické metody.

4. **Exploratorní analýza kategoriálních dat**. Popište metody jednorozměrné a mnohorozměrné exploratorní analýzy pro kategoriální proměnné. Dále se zaměřte na analýzu kontingenčních tabulek. Popište znaménkové schéma. K čemu se používá? Co to je a k čemu se používá korespondenční analýza? Popište její základní myšlenky. Jak se interpretují její výsledky?

5. **Shluková analýza**. Popište úlohu shlukové analýzy. Popište algoritmus a uveďte metody hierarchického shlukování. V čem se nehierarchické shlukování liší od hierarchického? Popište metodu k-means a k-medoids. Jaké metody se používají pro určení výsledného počtu shluků?

6. **Analýza nákupního košíku**. Popište analýzu nákupního košíku. Jaké číselné charakteristiky pravidel se používají? Jak se hledají pravidla pro dvou i víceprvkové množiny (apiori algoritmus)? Popište její zobecnění pro negované položky a hierarchické struktury dat.

7. **Lineární regrese**. Popište model lineární regrese, jeho předpoklady a interpretujte parametry modelu. Jaké metody se využívají pro výběr finálního modelu? Co to je multikolinearita? Jak se identifikuje a jaké může mít následky? Popište hřebenovou regresi a LASSO. K čemu se tyto metody používají?

8. **Logistická regrese**. Popište model logistické regrese. Jak se interpretují jednotlivé parametry tohoto modelu? Co to je logistické skóre? Jak se v logistické regresi odhadují hodnoty závisle proměnné? Co to je ROC a Lorenzova křivka? Uveďte číselné charakteristiky odvozené od těchto křivek.

9. **Rozhodovací stromy**. Jakou úlohu řešíme pomocí rozhodovacích stromů? Podrobně popište algoritmy CART a CHAID. K čemu slouží a jak funguje prořezávání? Uveďte číselné charakteristiky popisující kvalitu modelu. Uveďte některá rozšíření a zobecnění rozhodovacích stromů.