**Aula Prática 1**

**Tema: Introdução a Mineração de Dados e Representação dos Dados**

**Murielly Oliveira Nascimento – 11921BSI222**

**Perguntas**

1. Responda às seguintes perguntas
   1. Defina o que você entende por mineração de dados.

Mineração de Dados se dedica ao estudo e análise de grandes conjuntos de dados, usando algoritmos complexos. Em grande parte, técnicas de IA e Machine Learning são empregadas para extração de informações.

* 1. Cite 2 exemplos de aplicações do mundo real que pode se beneficiar do uso de técnicas de mineração de dados.

É possível com o auxílio de técnicas de Mineração de Dados identificar pacientes com COVID-19 ou câncer de mama e outras doenças. Também é possível, com base em valores de ações históricos, fazer previsões sobre a queda ou alta da bolsa de valores.

* 1. Porque você acredita que seja importante estudar mineração de dados. Como isso pode ser útil na profissão ou no seu trabalho de pós-graduação?

O fenômeno do Big Data ganhou ênfase nos últimos anos. Cada vez mais se discute as possibilidades e informações que podem ser extraídas em grandes conjuntos de dados. Inclusive para dados não estruturados (textos, posts, tweets etc.). O uso de técnicas de Mineração de Dados na Bolsa de Valores me chama atenção. Encontrar padrões a partir de séries históricas de ações e índices econômicos pode nos dar *insights* sobre a economia brasileira.

* 1. Procure na internet e cite o nome de pelo menos 2 algoritmos de classificação de dados e dois algoritmos de agrupamento de dados.

Algoritmos de classificação: Árvore de Decisão e Naive Bayes

Algoritmos de Agrupamento: K-médias e K-Medoids

1. Implementação sobre técnicas de amostragem
   1. Implemente a técnica de amostragem sem reposição vista em sala de aula, em que uma vez que uma instância foi selecionada para a amostra, ela é retirada da população. Essa técnica deve funcionar para qualquer base de dados, independente do número de instância e números de atributos. Escreva primeiramente o pseudo-código e discuta com o professor. A seguir implemente usando sua linguagem de programação de preferência.
   2. O seu programa deve receber como entrada um arquivo.csv e o tamanho da amostra a ser criada. O programa deve produzir como saída um arquivo.csv com as instâncias da amostra selecionada, uma em cada linha.
   3. Busque por uma base de dados no formato .csv em um repositório de bases de dados públicas. Por exemplo <https://archive.ics.uci.edu/ml/datasets.php>.
   4. Execute a técnica implementada usando a base de dados criada no item a.