**Bankruptcy Prediction Using Machine Learning: A Detailed Analysis**

**Executive Summary**

This project explores the potential of machine learning for predicting bankruptcy using financial data. Two popular algorithms, K-Nearest Neighbors (KNN) and Decision Tree, are implemented and evaluated for their accuracy in classifying companies as bankrupt or non-bankrupt. The project demonstrates the feasibility of utilizing machine learning for this task, offering valuable insights for financial institutions and investors in managing risk.

**1. Introduction**

Financial distress and bankruptcy pose significant challenges for financial institutions and investors, leading to substantial losses and hindering economic growth. Accurately predicting the risk of bankruptcy is crucial for making informed investment decisions and allocating resources efficiently. Traditional statistical methods have limitations in capturing complex relationships within financial data, leading to less than optimal predictions. Machine learning offers a promising alternative, capable of learning complex patterns and making more accurate predictions.

**2. Data Preprocessing**

This project utilizes a publicly available dataset containing financial information for a set of companies. The data is preprocessed using the Pandas library to address missing values, inconsistencies, and irrelevant features. Descriptive statistics are computed to understand the data distribution and identify potential outliers.

**2.1 Data Exploration**

Visualization techniques, such as swarmplots, are employed to explore the relationship between individual features and the target variable, "Bankrupt?". This analysis helps identify features that are most informative for predicting bankruptcy.

**2.2 Feature Engineering**

Highly correlated features are identified and removed to avoid overfitting and improve model performance. This process ensures that the model learns from independent pieces of information.

**3. Model Development**

**3.1 K-Nearest Neighbors (KNN)**

KNN is a simple yet powerful algorithm that classifies data points based on the majority class of their nearest neighbors. The "n\_neighbors" parameter is tuned to find the optimal value that balances accuracy and computational efficiency.

**3.2 Decision Tree**

Decision trees are non-linear models that learn decision rules by recursively splitting the data based on the most informative features. The "max\_depth" parameter is tuned to control the complexity of the tree and prevent overfitting.

**4. Model Evaluation**

The dataset is split into training and testing sets to train and evaluate the models. The performance of the models is assessed using classification reports, which include metrics such as accuracy, precision, recall, and F1-score. These metrics provide insights into how well the models correctly classify bankrupt and non-bankrupt companies.

**5. Results and Discussion**

The evaluation results demonstrate that both KNN and Decision Tree models achieve promising performance in predicting bankruptcy. The KNN model achieves higher accuracy, while the Decision Tree model exhibits better precision and recall. The choice of the optimal model depends on the specific needs and priorities of the user.

**5.1 Advantages and Limitations of KNN**

Advantages of KNN include its simplicity, interpretability, and ability to handle complex data relationships. However, KNN can be computationally expensive for large datasets and sensitive to the choice of the "n\_neighbors" parameter.

**5.2 Advantages and Limitations of Decision Tree**

Decision trees offer clear decision rules for understanding the model's predictions. However, they can be prone to overfitting and may not capture complex relationships in the data.

**6. Conclusion and Future Work**

This project demonstrates the promising potential of machine learning for bankruptcy prediction. Both KNN and Decision Tree models achieved significant accuracy in classifying companies as bankrupt or non-bankrupt. Future work could investigate the following:

* Experimenting with other machine learning algorithms, such as Support Vector Machines or Random Forests.
* Refining the data preprocessing techniques to improve the quality of the data.
* Exploring feature engineering techniques to create more informative features.
* Investigating the explain ability of the models to gain deeper insights into the factors driving the predictions.
* Developing a web application or API that allows users to leverage the trained models for real-time bankruptcy prediction.