PROJECT 1

Milestone 1- Rojan Khatri

Topic:

"Predicting Startup Funding Success- A Machine Learning Approach"

For this project, a predictive model will be developed to estimate the likelihood of startup companies receiving funding based on location, industry, and funding history.

Business Problem:

The project will address the uncertainty surrounding the startup funding challenges by predicting the probability of a startup securing investment. This will assist both investors and startups in making informed decisions, contributing to the overall success and growth of the startup ecosystem.

Datasets:

The primary dataset for this project is the *'Startup Investments Crunchbase'* from Kaggle. This dataset contains comprehensive information about startup companies, funding rounds, investors, and other related details from the Crunchbase platform. Besides, it also encompasses location, industry, funding history, and other relevant features crucial for predicting funding success.

Methods:

The project will utilize binary classification models to predict whether a startup will likely receive funding. Logistic regression and decision tree models will be used due to their suitability for binary classification tasks. In the beginning, the exploratory data analysis (EDA) will also be conducted to understand patterns and relationships within the data. Feature engineering will be applied to enhance model performance.

Ethical Considerations:

Potential ethical concerns include biases in the dataset, especially if certain groups or regions are overrepresented. To address this, the data will be thoroughly reviewed for biases, and adjustments will be made, if necessary. Additionally, transparency in model decision-making

will be prioritized to ensure fairness and unbiased predictions.

Challenges/Issues:

Challenges may include potential biases in the dataset, data quality issues, and the need for accurate predictive features. Addressing these challenges will involve careful data preprocessing, thorough validation of results, and, if necessary, seeking additional data sources to enhance the analysis.

This proposal provides a concise overview of the project, outlining its objective, the business problem it addresses, the selected dataset, analysis methods, ethical considerations, potential challenges, and references for validation.

References:

Özkurt, C. (2022). "Startup Investments EDA." Kaggle. Retrieved [December 3rd, 2023] from https://www.kaggle.com/code/cihatzkurt/startup-investments-eda

Andy_M. (2020, February 17). Startup investments (crunchbase). Kaggle. https://www.kaggle.com/datasets/arindam235/startup-investments-crunchbase?resource=download