1. Introduction

1.1 Overview

Startups are the backbone of modern innovation, yet nearly 90% fail within their first few years. This project aims to address the high risk of failure by developing a Machine Learning model that predicts whether a startup will eventually be **Acquired** or **Closed**. By leveraging historical data and identifying critical performance indicators, this tool provides entrepreneurs and investors with data-driven insights to mitigate risk.

1.2 Motivation

The motivation stems from the need to reduce the high rate of business disappointment caused by poor strategic planning and inadequate asset utilization. While human intuition is valuable, it is often subject to cognitive biases. An AI-driven approach offers an objective framework to assess a venture's probability of success long before the execution cycle is complete.

1.3 Problem Definition

To build a reliable classification system that:

- Analyzes 9 key startup parameters.
- Predicts the probability of a startup being acquired vs. closed.
- Utilizes a Random Forest algorithm to achieve high predictive accuracy.

---

2. Project Design

2.1 Requirements

- **Hardware**: Intel i5 processor or above, minimum 8GB RAM.
- **Software**: Python 3.x, Flask (for deployment), and libraries including Pandas, Scikit-learn, and Joblib.

2.2 Dataset Design

The model utilizes a dataset (e.g., from Kaggle) containing thousands of historical startup records. Key features include:

1. **Funding Total (USD)**: Total capital raised.
2. **Milestones**: Number of significant business achievements reached.
3. **Relationships**: Number of professional connections and partnerships.
4. **Funding Rounds**: Frequency of investment cycles.
5. **Geographic & Sector data**: State, city, and industry category.

---

3. Implementation & Analysis

3.1 Random Forest Algorithm

The project employs a **Random Forest Classifier**, an ensemble learning method that constructs multiple decision trees during training.

- **Mechanism**: It operates by taking the "majority vote" of all trees to determine the final prediction.
- **Benefit**: This approach is highly effective for handling complex, non-linear relationships in startup data and is more robust against overfitting than single decision trees.

3.2 Model Performance

- **Accuracy**: The model achieves a verified **80% accuracy** rate in classifying startup outcomes.
- **Evaluation Metrics**: Performance was measured using a confusion matrix, precision, and recall to ensure balanced detection of both successful and failed ventures.

---

4. Results & Conclusion

4.1 Screen Proofs

- **Input Interface**: A professional 9-field form designed for comprehensive data collection.
- **Prediction Outcome**: The system successfully differentiates between "CLOSED" and "ACQUIRED" statuses based on real-time user input.

4.2 Conclusion

This project successfully integrates a high-accuracy Machine Learning model with a modern web interface. The **Random Forest** algorithm proved to be the most effective for this

classification task, providing a stable and reliable tool for predicting startup viability. Future enhancements could include adding real-time social media sentiment analysis and founder experience profiles to further refine accuracy.