# Progress Report

## Project Title: Bankruptcy Prediction for Polish Companies

The project aims to predict the likelihood of bankruptcy in Polish companies using advanced machine learning techniques. Our methodology involved systematic steps, including data preprocessing, feature selection, model training, hyperparameter tuning, and implementation of ensemble learning techniques. The final optimized model utilized a Hybrid Stacking Ensemble approach, achieving outstanding predictive performance.

## Key Achievements

### 1. Model Implementation

- We implemented several machine learning models, including Logistic Regression, Random Forest Classifier, XGBoost, and Neural Networks (trained using a Stacked Autoencoder (SAE)).

- A Hybrid Stacking Ensemble model combining Neural Network and XGBoost predictions was built with Logistic Regression as the meta-learner.

### 2. Hyperparameter Tuning

- Keras Tuner was used for optimizing the architecture of the Neural Network.

- Optuna and GridSearchCV were utilized for fine-tuning XGBoost hyperparameters, ensuring maximum performance.

### 3. Final Model Performance

- The final Hybrid Stacking Model (Optimized Neural Network + XGBoost) achieved a remarkable accuracy of 97% on the test dataset.

Detailed performance metrics are as follows:

|  |  |  |
| --- | --- | --- |
| Metric | Class 0 (Non-Bankrupt) | Class 1 (Bankrupt) |
| Precision | 0.98 | 0.69 |
| Recall | 0.98 | 0.71 |
| F1-Score | 0.98 | 0.70 |

- Overall Accuracy: 97%

- Macro Average:

- Precision: 0.84

- Recall: 0.85

- F1-Score: 0.84

- Weighted Average:

- Precision: 0.97

- Recall: 0.97

- F1-Score: 0.97

### 4. Insights

- The hybrid stacking approach effectively combined the strengths of Neural Networks and XGBoost to improve the overall predictive performance, particularly for the minority class (bankrupt companies).

- While the recall for bankrupt companies (Class 1) reached 71%, further steps could be explored to enhance the precision and recall balance for minority class predictions.

## Next Steps

### 1. Further Model Refinement

- Experiment with additional ensemble techniques to boost precision and recall for the minority class.

- Investigate alternative sampling methods to improve class balance further.

### 2. Feature Analysis

- Conduct deeper analysis on feature importance to understand the drivers behind bankruptcy predictions.

### 3. Model Deployment

- Prepare the final model for integration into a production environment.

- Develop a user-friendly interface for end-users to input financial data and receive predictions.

### 4. Documentation and Presentation

- Finalize project documentation, including methodology, results, and insights.

- Prepare a detailed presentation for stakeholders, highlighting the success of the Hybrid Stacking Model and its practical implications.

## Conclusion

The Hybrid Stacking Model (Optimized Neural Network + XGBoost) has demonstrated exceptional accuracy and robustness in predicting bankruptcy. This model’s performance highlights the value of ensemble learning and hyperparameter optimization in solving complex financial prediction problems.