Predicting Corporate Bankruptcy with Machine Learning: Integrating Financial and Macroeconomic Indicators - Evidence from U.S. Public Companies (1999–2018)

by

Azizbek Ussenov

Submitted to

Central European University

Department of Economics and Business

In partial fulfilment of the requirements for the degree of

Master of Economic Policy in Global Markets

Supervisors: Robert Lieli and Zsofia Barany

Vienna, Austria

2025

AUTHOR'S DECLARATION

I, Azizbek Ussenov, the undersigned candidate for the MA/MSc degree in *Economic Policy in Global Markets* declare herewith that the present thesis titled "Predicting Corporate Bankruptcy with Machine Learning: Integrating Financial and Macroeconomic Indicators - Evidence from U.S. Public Companies (1999–2018)" is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography.

I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright.

I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Vienna, on 09.06.2025

Signature:

Full Name: Azizbek Ussenov

Abstract

In this thesis, I predict the likelihood of a company's bankruptcy using macroeconomic indicators for American companies listed on the New York Stock Exchange and NASDAQ for the years between 1999 and 2018. Advanced machine learning models, such as XGBoost, are employed to compare them with traditional ones, like Logistic Regression. I implement a rolling window approach to consider time-dependent changes in both firm and economy-level conditions when predicting default. The findings demonstrate that the XGBoost model achieves higher accuracy compared to the logit model, with an pooled AUC score of 0.9443. Even though macroeconomic indicators - such as interest rates, inflation, and real GDP growth - add predictive power to the models, the contribution is not at a significant level. The potential reason can be attributed to the use of fixed and the same macro variables for companies, which fail to provide cross-sectional discriminative power in predicting bankruptcy. Overall, the results demonstrate the advantages of XGBoost in developing data-driven solutions for monitoring the financial health of companies.

Keywords: rolling window length, XGBoost, Logistic Regression, AUC score

Acknowledgments

I would like to extend my gratitude to Professor Robert Lieli and Zsofia Barany for their invaluable advice and guidance on the setup and structure of the thesis. Their support and feedback were immensely helpful in guiding me in the right direction to produce a high-quality thesis.

Table of Contents

List of Figures

List of Tables

1. Introduction

The bankruptcy of companies is a critical event that can have far-reaching effects for all parties involved, ranging from investors to employees. Although bankruptcy can sometimes be economically efficient for some companies, it can lead to significant financial losses when it occurs unexpectedly. It is, therefore, crucial to have accurate prediction models to mitigate financial risks and save the business from exiting the market. Even though it is a challenging task to perfectly predict bankruptcy due to volatile markets and economic shocks, early detection of bankruptcy signs can serve as the best tool for stakeholders to adjust business goals and avoid significant losses that could occur without prediction. In this way, companies can evaluate their financial health and make well-informed decisions.

In general, most studies discuss the logistic regression (logit) model as a core model for predicting bankruptcy. For example, Beaver et al. (2005) demonstrated that the logit model remains a benchmark for corporate bankruptcy prediction due to its simplicity in interpreting the coefficients and results. However, the problem is that logit uses a linear combination of the predictors in modeling the conditional probability of a binary outcome. Therefore, the model may be prone to multicollinearity issues, which are common in complex financial relationships in real-world applications. The next issue with the model is its reliance on firm-specific financial indicators while neglecting macroeconomic indicators, which are crucial for the broader economic health of the country. Therefore, there is a need for more complex models that can capture both linear and non-linear complex relationships in the data. Therefore, advanced machine learning (ML) models, such as extreme gradient boosting (XGBoost), outperform traditional models like logit and narrow the gap in addressing complex non-linear relationships that the latter cannot handle (Máté et al., 2023).

Despite the increased use of ML models with more accurate predictions, limitations still exist in the literature regarding the setup. A significant shortcoming in bankruptcy prediction models is the underutilization of macroeconomic variables. Most studies rely on firm-specific financial indicators, such as profitability, when developing prediction models and overlook economic indicators, such as GDP growth, interest rates, or inflation. A study by Lombardo et al. (2024) based its models on only firm-level financial indicators, without any financial ratios or macro variables, resulting in lower prediction accuracy scores for the most powerful models, such as XGBoost. This approach limits the real-world applicability of the models and negatively affects their robustness when significant economic cycles and shocks occur in practice. For example, Sousa et al. (2022) demonstrated that the inclusion of macro indicators in the model can increase the prediction accuracy by at least 10%. Therefore, they argue that relying solely on firm-level indicators may overlook the impact of the general economic situation on bankruptcy. Tinoco and Wilson (2013) further support this approach by revealing the models' failure to capture the systemic risks (instability due to broader economic factors). In other words, models that exclude macroeconomic variables fail to accurately reflect the actual financial performance of companies across different economic cycles or shocks. Furthermore, Curi and Bortolotti (2025) illustrated the superiority of models incorporating macro indicators over those solely using financial indicators, concluding that the macroeconomic context is important for prediction tasks in the context of Italian small and medium enterprises.

Another issue overlooked in corporate default prediction is the underutilization of temporal changes. In other words, most models take the data as static, ignoring the periods that are crucial in the financial world. This is because risks evolve for firms due to various reasons, such as the negative consequences of the Global Financial Crisis (GFC) having long-term effects on company performance. Therefore, it is better to use different strategies for window length, such as a rolling window length, to account for the historical effect of bankruptcy. The rolling window length could be the most effective tool in predictive models, particularly when macroeconomic variables are added to the data due to its time dimension. Additionally, relying on outdated data may not be the most effective strategy for corporate default tasks, given the evolving financial landscape. For example, Franch et al. (2022) discuss the limitations in the use of temporal effects, particularly those roughed up by events like the GFC in the prediction. This helps to understand the risks that firms can take to avoid any adverse consequences for their performance. For this reason, a study led by Inoue et al. (2016) emphasizes the importance of implementing the rolling window length technique when addressing time-varying problems, such as company failure, which is heavily reliant on historical data.

The last but not least issue in the literature is related to the use of original features in the data for prediction. The problem is that most datasets provide raw variables, such as net income, net sales, or cost of goods sold, which are important for assessing the financial health of a company. However, financial ratios like return on assets are significantly important in providing early signs of financial performance. For example, Lombardo et al. (2024) did not engineer financial ratios from raw indicators, resulting in lower scores for models like XGBoost, which could perform better with well-engineered features because it can capture complex relationships in the data. Barboza et al. (2017) indicate the significance of accurate feature selection and engineering over model choice. So, engineered features are tailored to the specific problem. In that case, the performance of the chosen, more powerful model can yield better results, as models cannot identify which features or interaction terms are most important. Therefore, it is essential for a corporate default prediction task to engineer relevant features that help models identify the complex relationships in the data.

These studies collectively demonstrate the importance of a holistic approach in accurately identifying the early signs of bankruptcy. Therefore, this thesis takes into consideration all the mentioned issues to increase the accuracy rate of the models. Notably, it attempts to engineer features relevant to assessing a company's performance, employs macroeconomic features to account for the effects of economic cycles, and addresses temporal issues by incorporating more periods for training the models. Therefore, I aim to answer the following research question "*How does the integration of macroeconomic indicators with firm-level financial data, using rolling windows of at least five years, improve the accuracy of corporate bankruptcy prediction for U.S. public companies?*" The following measures are taken to answer the research question:

1. Set up a model that predicts one year ahead of company bankruptcy by incorporating financial and macroeconomic indicators. This study will utilize American public companies, with financial indicators listed on the New York Stock Exchange and NASDAQ, and macroeconomic data from FRED. The time horizon is from 1999 to 2018.
2. Implement benchmark models, such as logit, and more advanced models, like XGBoost, to handle complex relationships in the data.
3. Apply a rolling window length to capture temporal context with a 5-year lag and compare the models.

The results of this thesis would be a valuable asset for policymakers and stakeholders to monitor financial stability through both firm-level and macroeconomic performances. One of the main highlights of this study is that the XGBoost model outperformed the traditional statistical model logit for all window lengths. I also demonstrate that the inclusion of macroeconomic indicators adds predictive power to the XGBoost model but not at a significant level as expected. The logit model cannot take advantage of macroeconomic variables due to additional multicollinearity introduced while engineering interaction terms. However, it is still helpful to integrate them with financial indicators to serve as early warning systems. Therefore, policymakers and stakeholders can monitor systemic risks to the financial performance of companies to avoid significant consequences brought about by economic cycles.

This thesis is structured as follows. Chapter 2 provides information on how data was collected and its general structure. Chapter 3 illustrates the setup for logit and XGBoost models. Chapter 4 demonstrates and discusses the results of the models across different window lengths. Chapter 5 concludes the thesis and presents policy implications.

2. Data

2.1 Firm-level data

The company-specific data is mainly confidential; therefore, this thesis utilizes a publicly available dataset comprising 8,971 United States public companies listed on the New York Stock Exchange and NASDAQ. The dataset was downloaded from the world's largest data science community, known as "Kaggle" (Singh, 2023). The period covered in the dataset spans from 1999 to 2018, yielding 78682 observations. In total, there are 21 features, of which 3 are company name, bankruptcy status, and year. The remaining features are firm-level financial indicators, which can be found in Table 1, sourced from Kaggle (Singh, 2023).

Table 1. The overview of the dataset about bankruptcy of American large companies

#	Variable	Description
X1	Current assets	All the assets of a company that are expected to be sold or used as a result of standard business operations over the next year.
X2	Cost of goods sold	The total amount a company paid as a cost directly related to the sale of products.
X3	Depreciation and amortization	Depreciation refers to the loss of value of a tangible fixed asset over time (such as property, machinery, buildings, and plant). Amortization refers to the loss of value of intangible assets over time.
X4	EBITDA	Earnings before interest, taxes, depreciation, and amortization. It is a measure of a company's overall financial performance, serving as an alternative to net income.
X5	Inventory	The accounting of items and raw materials that a company either uses in production or sells.
X6	Net Income	The overall profitability of a company after all expenses and costs have been deducted from total revenue.
X7	Total Receivables	The balance of money due to a firm for goods or services delivered or used but not yet paid for by customers.
X8	Market value	The price of an asset in a marketplace. In this dataset, it refers to the market capitalization since companies are publicly traded in the stock market.
X9	Net sales	The sum of a company's gross sales minus its returns, allowances, and discounts.
X10	Total assets	All the assets, or items of value, a business owns.

X11	Total Long-term debt	A company's loans and other liabilities that will not become due within one year of the balance sheet date.
X12	EBIT	Earnings before interest and taxes.
X13	Gross Profit	The profit a business makes after subtracting all the costs that are related to manufacturing and selling its products or services.
X14	Total Current Liabilities	The sum of accounts payable, accrued liabilities, and taxes such as Bonds payable at the end of the year, salaries, and commissions remaining.
X15	Retained Earnings	The amount of profit a company has left over after paying all its direct costs, indirect costs, income taxes, and its dividends to shareholders.
X16	Total Revenue	The amount of income that a business has made from all sales before subtracting expenses. It may include interest and dividends from investments.
X17	Total Liabilities	The combined debts and obligations that the company owes to outside parties.
X18	Total Operating Expenses	The expenses a business incurs through its normal business operations.

2.2 Macro-level data

The macroeconomic variables were downloaded from the Federal Reserve Economic Data (FRED) database. All variables have a yearly frequency for the years from 1999 to 2018. So, they are fixed for each company and year and concatenated with financial indicators accordingly. The list of variables used is presented in Table 2.

Table 2. The macroeconomic indicators

#	Variable	Description
Z1	Real GDP Growth	The rate of inflation-adjusted economic output over time.
Z2	Inflation	The rate of the general level of prices for goods and services rises and erodes the purchasing power.
Z3	Interest Rate	The Federal Funds Effective Rate
Z4	Unemployment Rate	The unemployment rate
Z5	Credit Spread	ICE BofA US High Yield Index Option-Adjusted Spread

3. Methodology

3.1 Data split and rolling window length

This thesis predicts one year ahead of company bankruptcy, using a dataset that spans 20 years. To address the discussed gaps in the literature regarding temporal and macroeconomic contexts, a rolling window length with a 5-year lag was implemented. Therefore, there are three cases for model comparison: the model using original features, the model using original features with financial ratios, and the model using everything with macroeconomic variables. The rolling window length strategy consists of 2 steps. The first step is to train the data on the specified date. Use a 3-year validation set to tune hyperparameters and predict the default for the following year. The second step is to roll the window by one year, meaning all train, validation, and test sets are rolled by one year. This strategy is repeated until the last year in the dataset. For comparison purposes, window lengths (wl) of 5 years and 9 years were chosen to incorporate sufficient historical data for model training. Wl = 5 comprised seven splits, and wl = 9 comprised three splits. The structure sample for wl is demonstrated in Table 3 and Table 4. The first year in the training set starts from 2004 because 5 years of lags implemented, so the year 2004 includes historical data from 1999, 2000, 2001, 2002, and 2003

Table 3. wl = 5 data structure

Split #	Train Set					Validation Set			Test Set
1	2004	2005	2006	2007	2008	2009	2010	2011	**2012**
. .	. .	. .	. .	. .	. .	. .	. .	. .	. .
7	2010	2011	2012	2013	2014	2015	2016	2017	**2018**

Table 4. wl = 9 data structure

Split #	Train Set					Validation Set			Test Set
1	2004	2005	…	2011	2012	2013	2014	2015	**2016**
2	2005	2006	…	2012	2013	2014	2015	2016	**2017**
3	2006	2007	…	2013	2014	2015	2016	2017	**2018**

There is a significant class imbalance, and Figure 1 and Figure 2 illustrate the bankrupt percentage per split for each window length. This issue should be taken into consideration when running models to avoid over-representation of the majority class.

Figure 1. The class imbalance per split for wl = 5

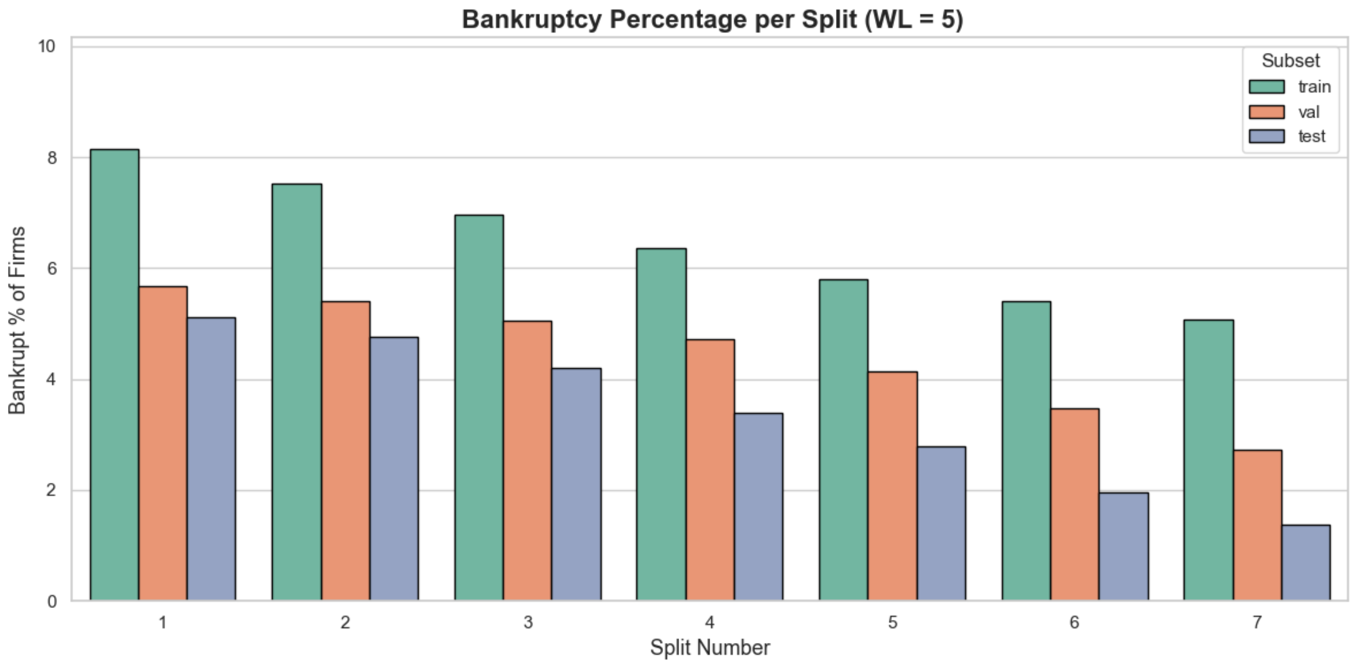
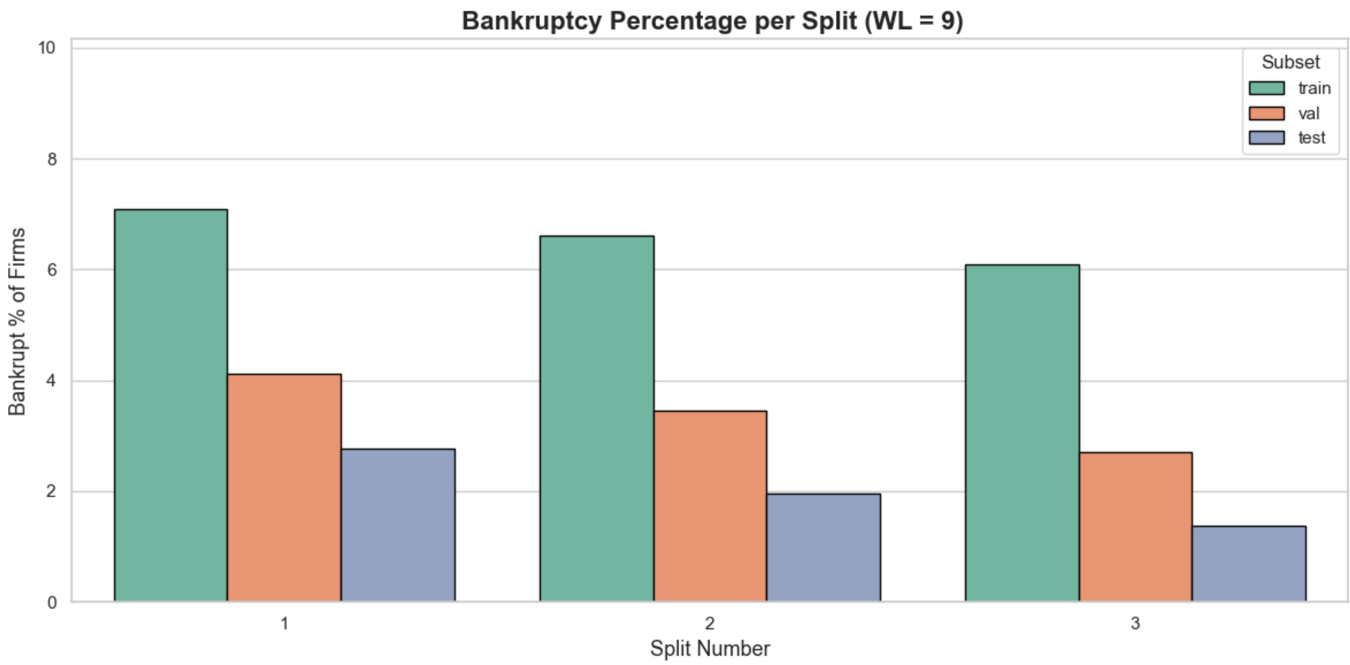


Figure 2. The class imbalance per split for wl = 9



3.2 Feature Engineering

As Barboza et al. (2017) pointed out regarding the importance of feature selection over model choice, careful selection was employed in this thesis. After several trials, 25 new financial ratios, such as return on assets, asset turnover, and gross profit to assets, were developed on top of the original ones. For example, return on assets refers to the amount of profit a company generates from its assets or the gross profit to assets refers to how efficiently a company utilizes its assets to generate a gross profit. These ratios can help assess the company's financial health more broadly.

For macroeconomic indicators, interaction terms with original features, such as total liabilities to interest rates or gross profit to inflation, and others, were developed. For example, total liabilities to interest rate means a high interest rate incurs more debt costs. Gross profit to inflation captures the effect of inflation on profit margin; that is, high inflation erodes profit margin, which is not a good case for companies.

3.3 Transformations

Real-life data often exhibit a complex distribution, and addressing skewness is crucial before running models to avoid biases and reduced predictive performance. This study employed the combination of Power Transformation with Robust Scaler. Power Transformation is helpful in addressing skewness in numerical features to avoid any long tails. Robust Scaler utilizes the median and interquartile ranges to account for outliers, which are common events in the financial world due to market volatility. These transformations were applied to numerical features; however, they were unable to handle the year data correctly. To correctly normalize the years in the data to avoid any large magnitudes across features, they were normalized to a [0, 1] range.

3.4 Machine Learning Models

3.4.1 Logistic Regression

Logistic Regression (logit) models the relationship between a binary dependent variable and one or more independent variables. The primary requirement of this model is to have low multicollinearity among the independent variables. The equation for logit is given by *Equation 1* and *Equation 2*.

Equation 1: $z = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_mx_m$

Equation 2: $P(Y = 1|x_1, \ldots, x_m) = \frac{1}{1+e^{-z}}$

where the x variable is a feature vector and z is just a linear representation of x vectors. The model predicts the likelihood of bankruptcy based on a set of features. Typically, if the likelihood is above 0.5, the company will fail (class 1); however, the threshold can be adjusted accordingly. (Máté et al., 2023).

I used random search for hyperparameter tuning (RandomizedSearchCV) due to its advantage in computational cost and scalability (Bergstra & Bengio, 2012). The predefined parameter values are passed to the random search, and the best ones are identified based on the area under the curve (AUC, explained in the following subsection) score for my task. I tuned the following parameters to avoid overfitting and account for the trade-off between bias and variance: regularization strength parameters C, L1, and L2.

3.4.2 XGBoost

The XGBoost machine learning model is a member of the gradient boosting technique family but with extreme efficiency and scalability. It is a tree-based model, meaning it builds an ensemble of trees, but each tree is constructed to minimize the errors made by the previous tree, thereby increasing accuracy. This model is powerful enough to handle multicollinearity, normality, and missing value issues. XGBoost sums up *K* functions to predict the outcome based on *Equation 3* (Chen & Guestrin, 2016).

Equation 3: $\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), f_k \in F$

where $F = \{f(x) = \omega_{q(x)}\}(q: R^m \to T, \omega \in R^t)$. So, T is the number of leaves in the tree, w is leaf weights, q is the structure of the tree that maps data to the leaf, and f_k is an independent tree structure with q and w.

The model minimizes the regularized (L1 and L2) objective function *L(ϕ)* given in *Equation 4* to shrink the difference between target and predicted values. Here is a penalty to avoid the complexity and overfitting in the model (Chen & Guestrin, 2016).

Equation 4: $\boldsymbol{L(\phi)} = \sum_{i} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$

where $\Omega(f) = \gamma T + \frac{1}{2}\lambda||\omega||^2$ and γ and λ are regularization parameters.

XGBoost utilizes a second-order Taylor expansion to optimize the loss function in *Equation 4* efficiently. In this way, the model can approximate the complete loss instead of evaluating it at each split. The output is shown in *Equation 5* (Chen & Guestrin, 2016).

Equation 5: $L^{(t)} = \sum_{i=1}^{n} [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t)$

where $g_i = d_{\hat{y}_i^{(t-1)}}\ l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = d^2_{\hat{y}_i^{(t-1)}}\ l(y_i, \hat{y}_i^{(t-1)})$

Since XGBoost is a powerful model, it utilizes numerous parameters within its algorithm. I utilized an advanced hyperparameter optimization framework called "Optuna" to fine-tune the model's parameters. Optuna is the best tool due to its use of neural network architecture in maximizing or minimizing the objective function (Akiba et al., 2019). The following functional parameters for corporate default tasks were tuned with Optuna to avoid overfitting, minimize model complexity, handle class imbalance, and improve generalization: maximum depth of the tree, number of estimators, learning rate, L1 regularization, L2 regularization, and weight for the positive class.

3.5 Evaluation Metrics

Predicting corporate bankruptcy is a binary classification task. In this thesis, there are two classes: Class 1 consists of bankrupt companies, and Class 0 consists of alive companies. The following variables are used to explain the evaluation metrics:

1. True Positive (TP) - the number of companies that have been predicted as bankrupt when they are bankrupted.
2. False Positive (FP) - the number of companies that have been predicted as bankrupt when they are alive.
3. True Negative (TN) - the number of companies that have been predicted as alive when they are alive.
4. False Negative (FN) - the number of companies that have been predicted as alive when they are bankrupted.

The most common evaluation metric for default prediction tasks is the Area Under Curve (AUC) score. The curve's name is Receiver Operating Characteristic (ROC), and it plots the True Positive Rate (TPR) against the False Positive Rate (FPR) provided in *Equations 6 and 7*.

Equation 6: $TPR = \frac{TP}{TP+FN}$

Equation 7: $FPR = \frac{FP}{FP+TN}$

The AUC score is simply the area under the ROC curve, which helps distinguish between the positive and negative classes by trying different thresholds for predicted probabilities (Lombardo et al., 2024). In other words, it does not require a typical threshold of 0.5 to define a positive class, but rather, it experiments with different thresholds within the range of 0 and 1. This thesis primarily focuses on the results of the AUC score, and the models are also tuned and evaluated based on that score.

Among other useful evaluation metrics are micro-F1, macro-F1, Type I error rates, and Type II error rates. Micro-F1 score means that it takes into consideration the contributions made by all classes to calculate the average metric. On the other hand, macro-F1 treats all classes equally, calculates the F-1 (*Equation 8*) score for each class, and then averages the metric. Regarding Type I and Type II error rates, they help us understand at what level the model is incorrectly classifying bankrupt cases (*Equation 9*). The negative costs of a Type II error are higher than those of a Type I error because companies can incur unexpected financial losses, whereas a Type I error prompts companies to take precautionary actions.

Equation 8: $F1\ score = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}}$

where $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$

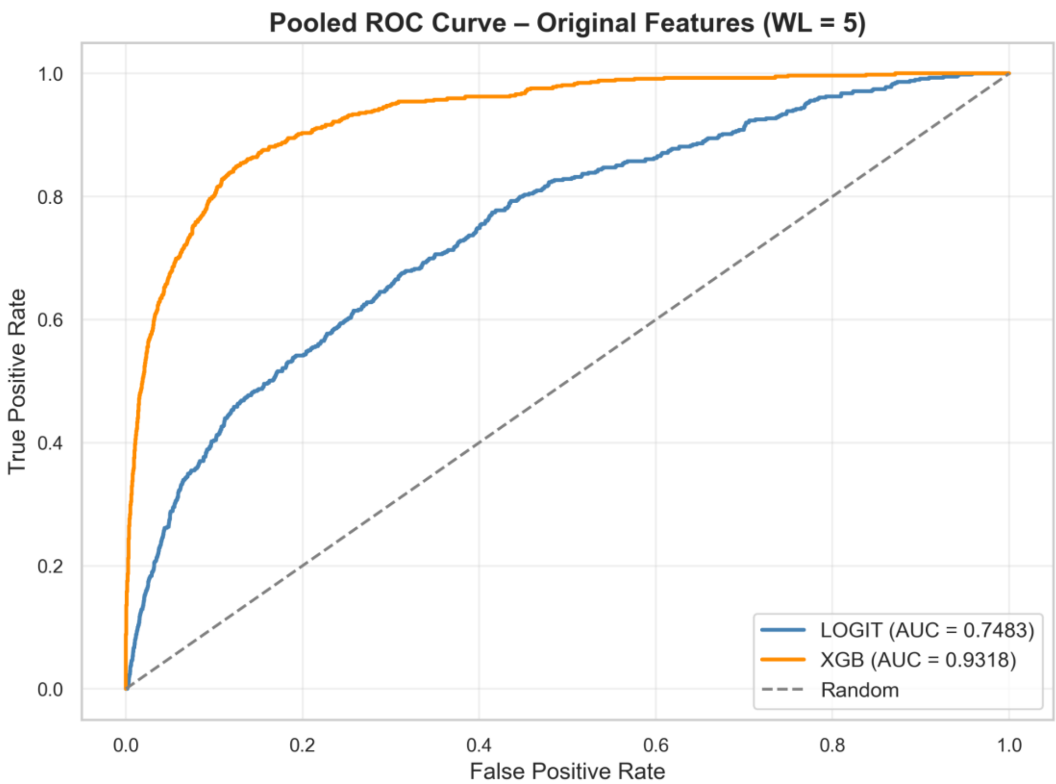
Equation 9: $Type\ I = \frac{FP}{TN+FP}$ and $Type\ II = \frac{FN}{TP+FN}$

4. Results and Discussion

There are two main sets of results: wl = 5 and wl = 9. First, the results for wl = 5 are presented, followed by those for wl = 9 for comparison purposes. Both logit and XGBoost models were implemented for wl = 5, but only the logit model was run for wl = 9 to understand the improvement when more historical data was used for the training set. Since XGBoost performed nearly flawlessly for wl = 5, there is no need to run it for wl = 9, as it is difficult to achieve an improvement in the AUC score beyond perfection.

The pooled AUC scores and plots across splits were computed due to the number of splits and three different cases with features. The results for wl = 5, when only the original features are used, are provided in *Figure 3*.

Figure 3. Pooled ROC Curve - Original Features for wl = 5



Here, the perfect model would be the one that touches the top-left corner, meaning TPR = 1 and FPR = 0. However, it is difficult to achieve in practice. The 45-degree line means the random guess model. In this figure, we can see that XGBoost achieved an AUC score of 0.9318 and a logit score of 0.7483 using only the original balance sheet features in the dataset. Both models significantly outperform a random guess, but XGBoost surpasses the logit model. The superiority of XGBoost means that it accurately distinguishes bankrupt companies from non-bankrupt ones with high TPR and low FPR. The finding supports the results from Lombardo et al. (2024), who used only original features for modeling, and XGBoost outperformed the logit.

Figure 4. Pooled ROC Curve - Original Features with Financial Ratios for wl = 5

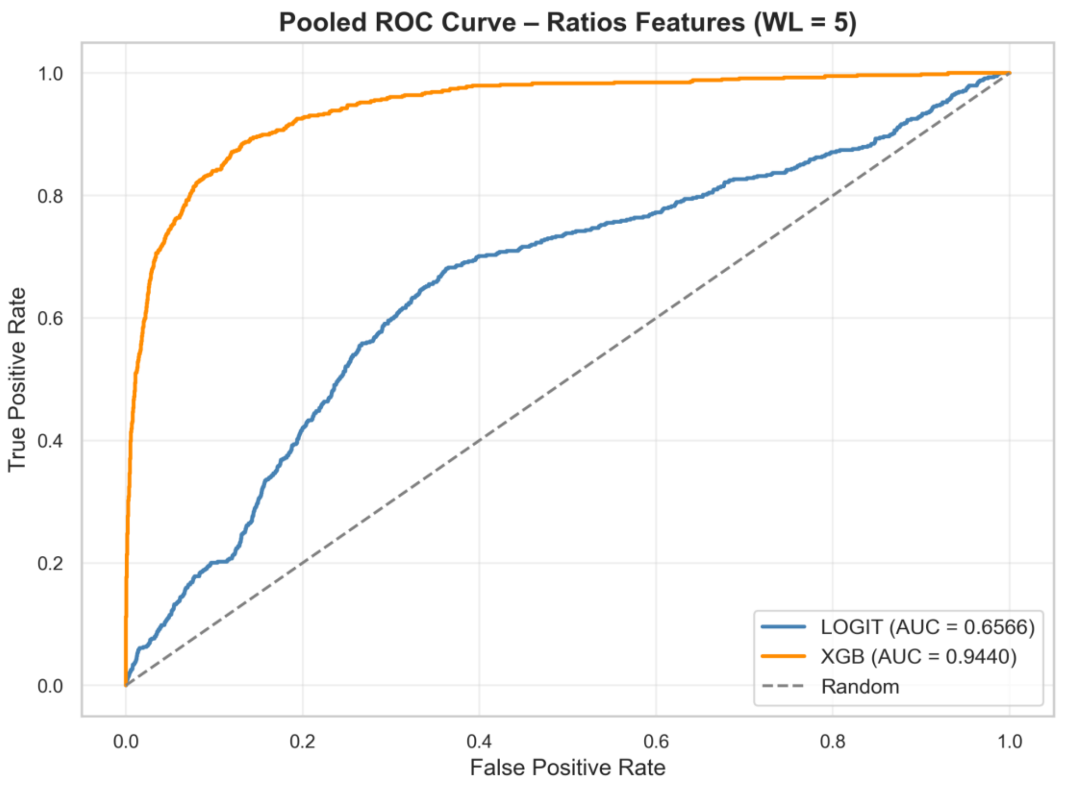
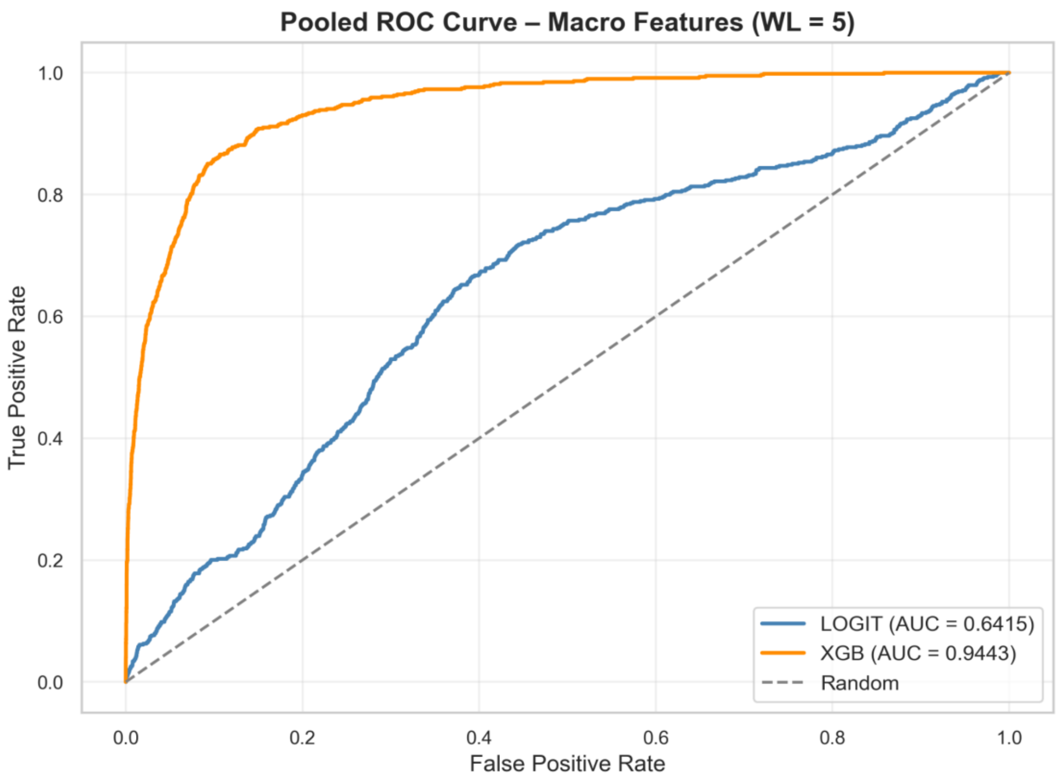


Figure 4 shows the results for the combination of original features with financial ratios. The XGBoost model benefited from engineered financial ratios, resulting in an increase in the AUC score from 0.9318 to 0.9440. However, the logit model's performance worsened because logit expects low multicollinearity, but the introduction of additional financial ratios caused multicollinearity, resulting in a decrease in performance. This means logit is struggling to distinguish the classes correctly.

The following figure presents the results for the model using macroeconomic indicators along with original and financial ratio features. The inclusion of macroeconomic variables contributed even more negatively to the logit model due to the additional multicollinearity introduced with the cross indicators. However, the XGBoost model was able to extract informative context from macro features and was used correctly in prediction, achieving an even higher AUC score of 0.9443. Therefore, it correctly distinguishes between bankrupt classes, aligning with the study's results by Curi and Bortolotti (2025). The addition of macro variables did not add significant predictive power, even for powerful models like XGBoost. The potential explanation could be the use of fixed macro features for all companies, as the economy is common for all companies in the US. In other words, the companies share common macro indicators that fail to provide cross-sectional discriminative power in prediction.

Figure 5. Pooled ROC Curve - everything with Macroeconomic Variables for wl = 5



The AUC scores for each split show the actual performance of the models for each year in the test set. Here, Figure *6* presents the trendline of scores across the splits for all feature and model cases; the actual AUC plots are reported in the Appendix. The figure illustrates the apparent dominance of the XGBoost model over the logit model for all splits and feature cases. The inclusion of macroeconomic indicators proved their importance in corporate prediction, resulting in a higher AUC score among all cases. The XGBoost model effectively handled the complex linear and non-linear relationships in the data, resulting in fewer fluctuations compared to the logit model. A potential reason for the significant fluctuations in the logit model is that the US economy faced considerable issues due to the GFC. The effect of this crisis lasted approximately up until 2015, corresponding to the split number 5. We can observe the increase in the performance of financial ratios and macroeconomic variables. A possible explanation is the reduced noise in the data and a more stable economy, which allows the model to capture the effect of macroeconomic variables accurately. In general, the logit model experienced a substantial increase in performance; however, it is still considerably lower than the XGBoost model.

Figure 6. Actual AUC scores for each model and split wl = 5

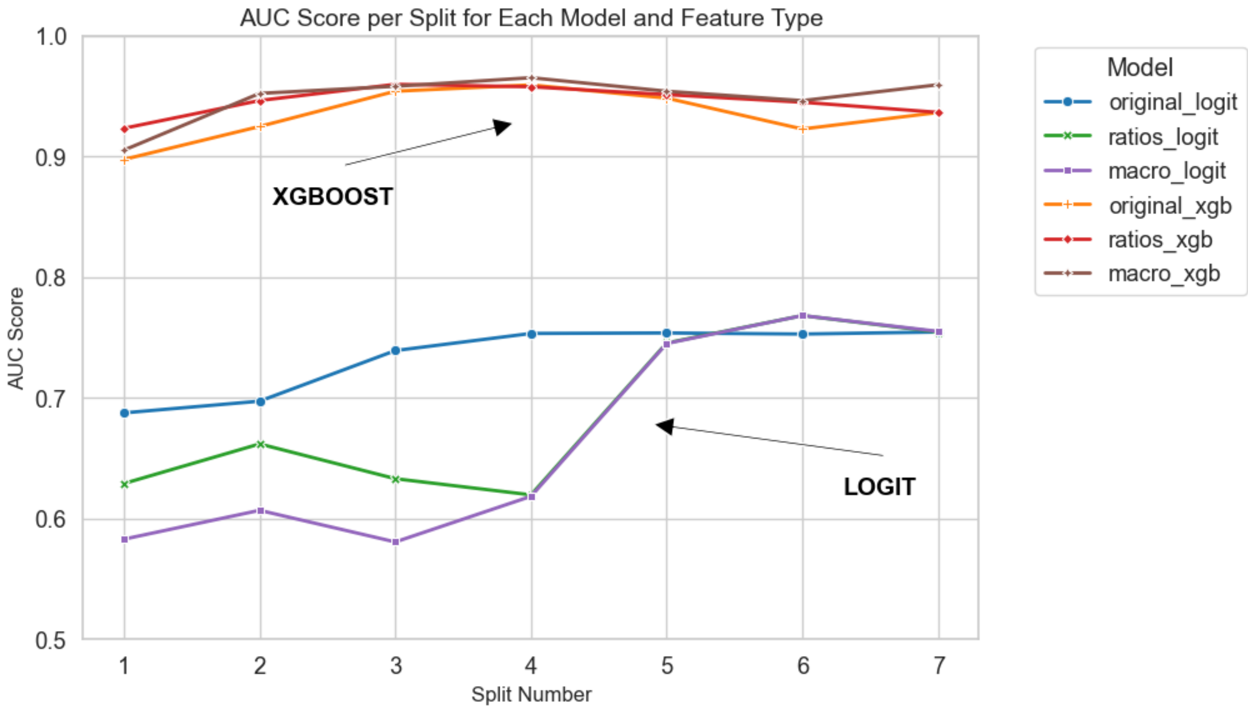


Table 5 below presents various evaluation metrics for the pooled performance of the logit model with wl = 5. As we can see, the AUC score of the logit model dropped when additional features were added due to multicollinearity issues. This can be explained by a high score of Type I errors, meaning that the model is facing serious issues in distinguishing between bankrupt and non-bankrupt cases. However, Type II errors indicate that the model is not missing as many bankrupt firms as bankrupt ones when making predictions. In addition, the overall drop in micro and macro-f1 scores means the model struggles with severe class imbalance and makes more mistakes.

Table 5. Pooled summary statistics for logit model: wl = 5

Model	TP	FN	FP	TN	AUC Score	Micro-f1	Macro-f1	I Error	II Error
original_logit	495	94	8749	7704	0.7483	0.4811	0.3680	0.5318	0.1596
ratios_logit	429	160	7850	8603	0.6566	0.5300	0.3895	0.4771	0.2716
macro_logit	462	127	9334	7119	0.6415	0.4448	0.3449	0.5673	0.2156

Table 6 presents the evaluation metrics for the powerful XGBoost model with wl = 5. The results are significantly better than those from the logit model. The AUC score, micro-F1 score, and macro-F1 score all experienced an increase, along with a drop in Type I errors, indicating improved performance in

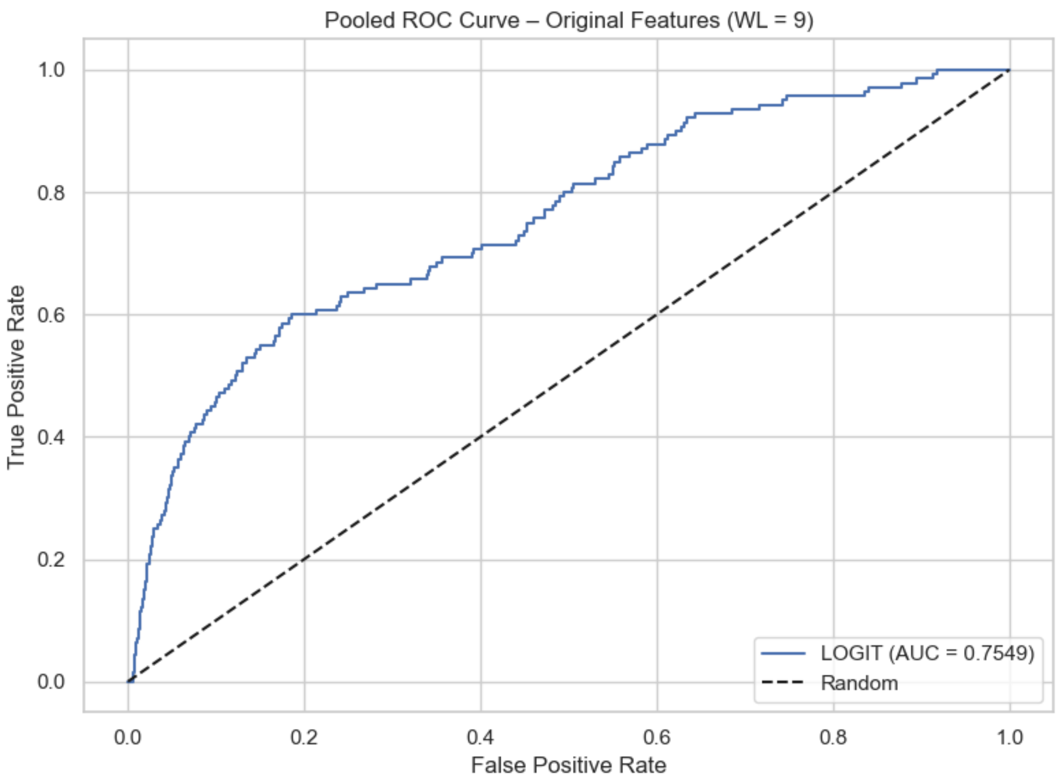
prediction. However, the high Type II errors mean that the model struggles to identify companies as bankrupt, often classifying them as alive. We can observe this event by comparing the high FN scores to the low scores using the logit.

Table 6. Pooled summary statistics for XGBoost model: wl = 5

Model	TP	FN	FP	TN	AUC Score	Micro-f1	Macro-f1	I Error	II Error
original_xgb	172	417	84	16369	0.9318	0.9706	0.6960	0.0051	0.7080
ratios_xgb	159	430	38	16415	0.9440	0.9725	0.6953	0.0023	0.7301
macro_xgb	172	417	68	16385	0.9443	0.9715	0.7002	0.0041	0.7080

The results for wl = 9 are provided in the figures below; however, only the logit model is presented, as the XGBoost performs almost perfectly for wl = 5. Therefore, only the logit was run to observe any change if more data was used for training. *Figure 7* shows the pooled AUC plot of the logit model using the original features. We can observe that the increased AUC score compared to wl = 5 results in fewer false positives.

Figure 7. Pooled ROC Curve - Original Features for wl = 9



The following pooled AUC plots, featuring financial ratios and macroeconomic variables, are presented in *Figures 8* and *9*, respectively. The scores are significantly higher than those from wl = 5, specifically increasing to 0.7572, but with more false positives. The model with financial ratios or all features, including macroeconomic ones, is better than the one using only original features. A possible explanation for this result is that including more historical data could help capture the long-term effects of financial and macroeconomic indicators. In other words, the indicators are more stabilized across a vast time horizon, so long-term effects contribute more to the model than short-term effects. The Appendix illustrates the individual AUC plot for each split.

Figure 8. Pooled ROC Curve - Original Features with Financial Ratios for wl = 9

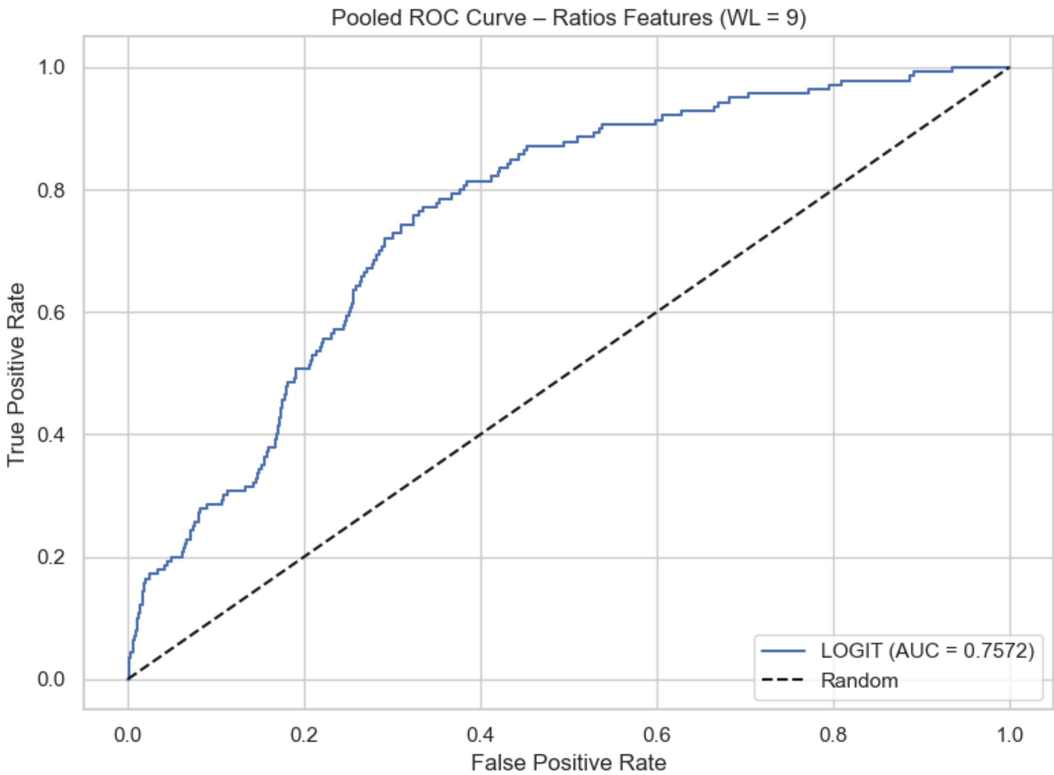
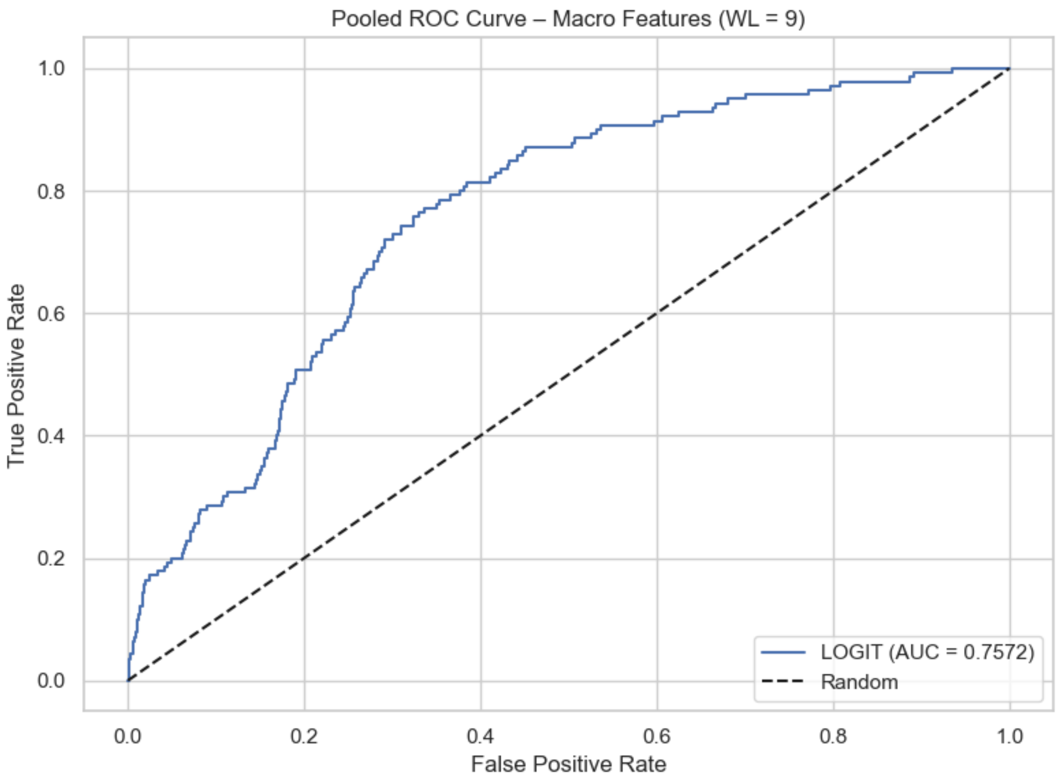


Figure 9. Pooled ROC Curve - everything with Macroeconomic Variables for wl = 9



The other evaluation metrics for wl = 9 are provided in *Table 5*. We can see a higher Type I error and a lower Type II error when including financial and macro features. Wl = 5 showed the opposite results. Overall, the results are in the opposite direction for wl = 9 compared to wl = 5, indicating that high values are replaced with low values or vice versa. The explanation could be the focus on more recent data, such as wl = 5, meaning more aggressive models but more false positives. When using a more extended period, such as wl = 9, the model becomes more conservative and attempts to produce fewer false positives.

Table 7. Pooled summary statistics for logit model: wl = 9

Model	TP	FN	FP	TN	AUC Score	Micro-f1	Macro-f1	I Error	II Error
original_logit	82	58	1217	5479	0.7549	0.8135	0.5049	0.1818	0.4143
ratios_logit	123	17	3345	3351	0.7572	0.5082	0.3671	0.4996	0.1214
macro_logit	122	18	3328	3368	0.7572	0.5105	0.3680	0.4970	0.1286

5. Conclusion

In this thesis, I developed company bankruptcy prediction models and compared the traditional logit model with more advanced machine learning models, such as XGBoost. I employed macroeconomic indicators in this prediction task but also used original features and financial ratios for comparison reasons. The setup for the model is crucial for obtaining accurate results. Therefore, the use of different rolling window lengths with 5 years of lags was chosen to utilize both more recent and longer historical data to predict defaults accurately. This setup was successfully implemented with macroeconomic features and resulted in higher performance for the XGBoost model for wl = 5. However, the logit model could not capture the complex relationships in the data for wl = 5, but it could handle them using wl = 9, leading to better performance. The overall result is that macroeconomic variables add predictive power for bankruptcy, but not at a significant level. Therefore, using balance sheet variables or financial ratios can also lead to high prediction accuracy.

This thesis has two policy implications for policymakers. Firstly, the accurate corporate default prediction serves as an early warning system for policymakers when using macroeconomic indicators in the modeling. Policymakers can adjust their policies in response to the macroeconomic context to avoid bankruptcies. Secondly, the use of more advanced machine learning models, such as XGBoost, is considered an efficient data-driven approach to developing more powerful predictive models. This thesis demonstrated the superiority of the XGBoost model in accurately predicting company defaults, allowing policymakers to modify the risk assessment framework to include XGBoost.

This thesis has a list of limitations. Firstly, I used only numeric data showing the financial health of the companies. However, it would also be beneficial to use textual data, such as news sentiments, to capture the current condition of companies along with financial indicators. Finally, the use of macroeconomic variables pertaining only to the US can not be transferable to other countries. This is because the economic conditions of other countries may significantly differ from those in the US. Therefore, it would be better to extend internationally to compare macro risks across countries.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. https://doi.org/10.1145/3292500.3330701

Barboza, F., Kimura, H., & Altman, E. I. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications, 83*, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

Beaver, W. H., McNichols, M. F., & Rhie, J.-W. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies*, *10*(1), 93–122. https://doi.org/10.1007/s11142-004-5861-1

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*, 281–305. http://www.jmlr.org/papers/v13/bergstra12a.html

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Curi, C., & Bortolotti, A. (2025). Predicting financial distress and bankruptcy in small and medium enterprises: Insights from machine learning models. *SSRN*. http://dx.doi.org/10.2139/ssrn.5236086

Federal Reserve Bank of St. Louis. (n.d.). *FRED, Federal Reserve Economic Data*.

Franch, F., Nocciola, L., & Vouldis, A. (2022). Temporal networks in the analysis of financial contagion. *ECB Working Paper No. 2667*. European Central Bank.

Inoue, A., Jin L., & Rossi, B. (2016). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics,* *196*(1), 55–67. https://doi.org/10.1016/j.jeconom.2016.03.006

Lombardo, G., Pellegrino, M., Adosoglou, G., Cagnoni, S., Pardalos, P. M., & Poggi, A. (2024). Machine learning for bankruptcy prediction in the American stock market: Dataset and benchmarks. *Future Internet, 14*(8), 244. https://doi.org/10.3390/fi14080244

Máté, D., Raza, H., & Ahmad, I. (2023). Comparative analysis of machine learning models for bankruptcy prediction in the context of Pakistani companies. *Risks,* *11*(10), 176. https://doi.org/10.3390/risks11100176

Singh, U. (2023). US company bankruptcy prediction dataset. *Kaggle*.

Sousa, A., Braga, A., & Cunha, J. (2022). Impact of macroeconomic indicators on bankruptcy prediction models: Case of the Portuguese construction sector. *Quantitative Finance and Economics, 6*(3), 405–432. https://doi.org/10.3934/QFE.2022018

Tinoco, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis, 30*, 394–419. https://doi.org/10.1016/j.irfa.2013.02.013

Appendix

The AUC plots for wl = 5 are provided between *Figure 1* and *Figure 21*.

Figure 1. ROC Curve - Split 1 - Original Features for wl = 5

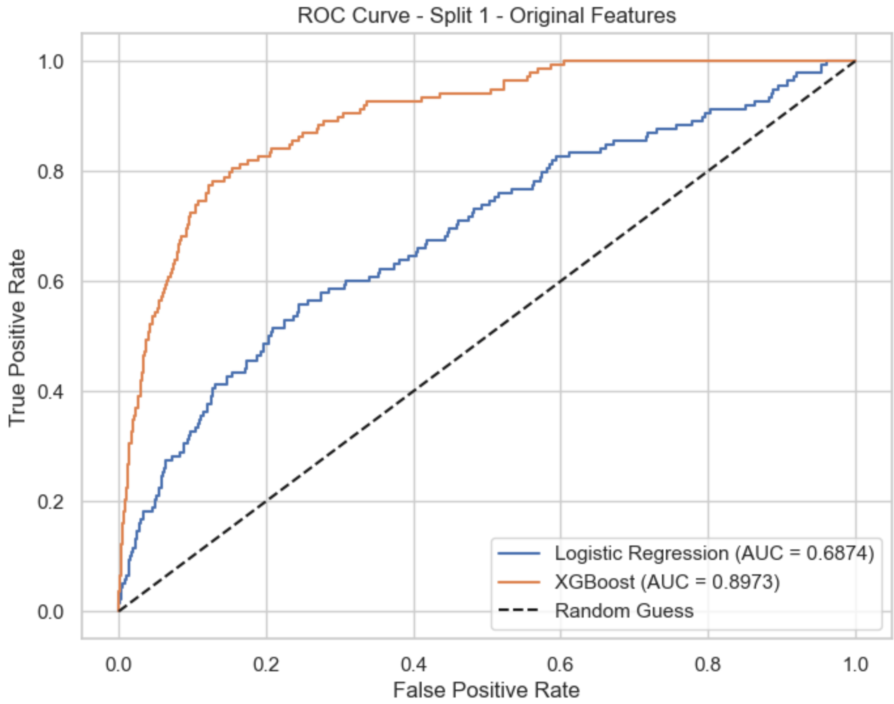


Figure 2. ROC Curve - Split 1 - Original Features with Financial Ratios for wl = 5

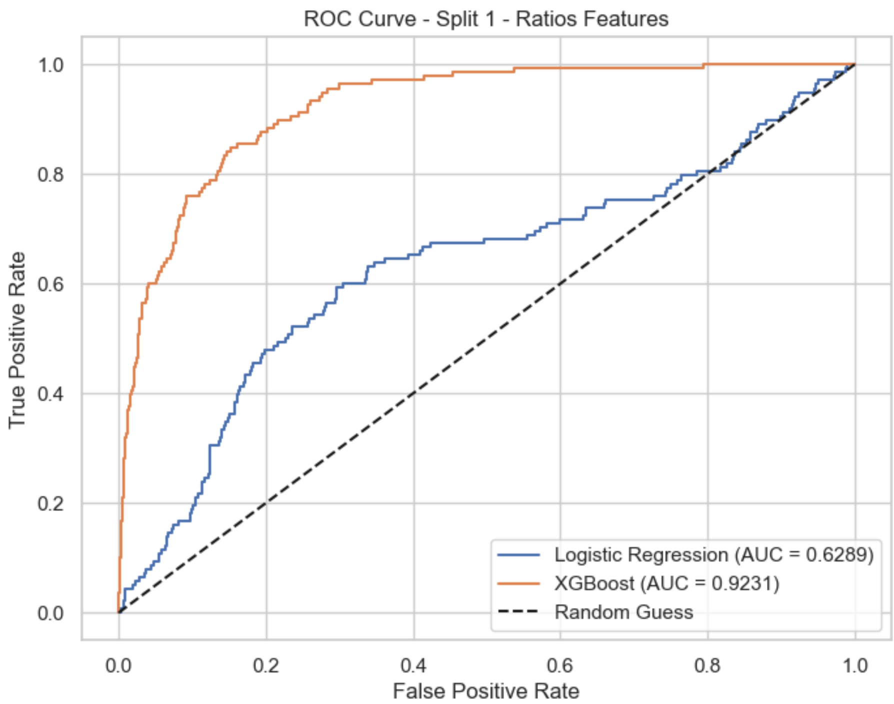


Figure 3. ROC Curve - Split 1 - everything with Macroeconomic Variables for wl = 5

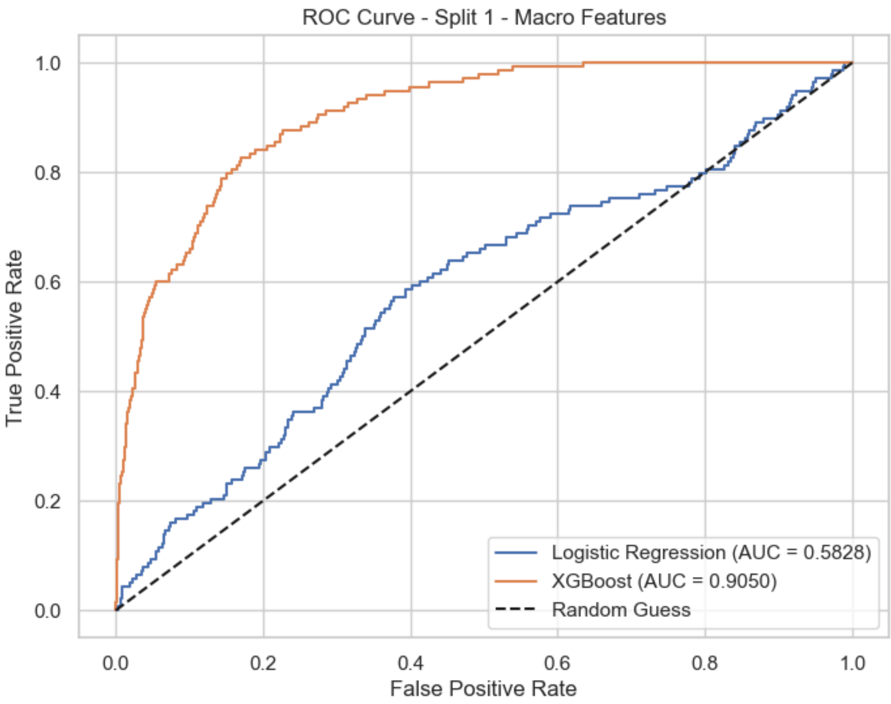


Figure 4. ROC Curve - Split 2 - Original Features for wl = 5

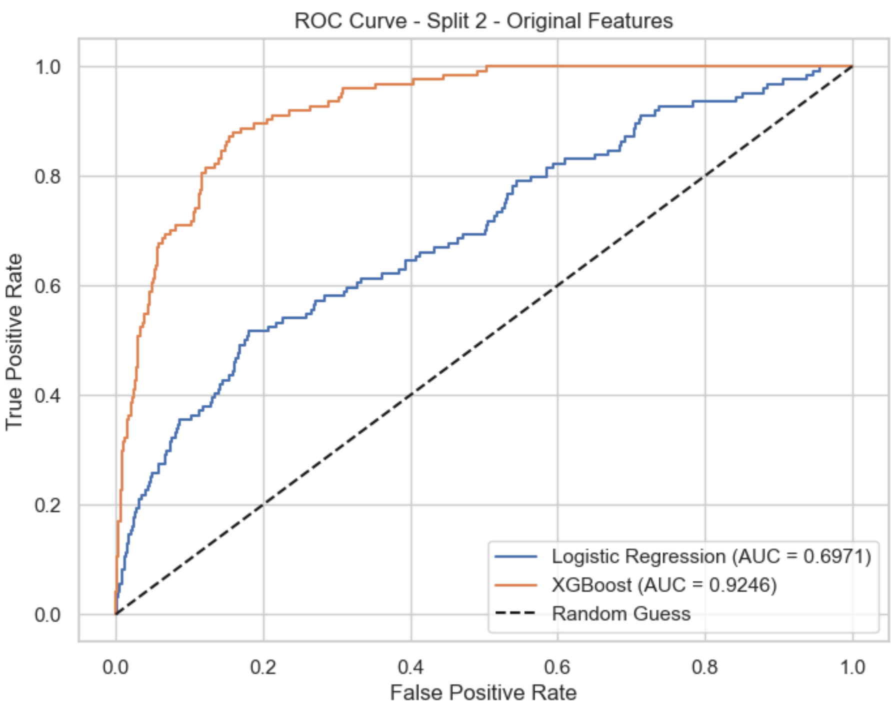


Figure 5. ROC Curve - Split 2 - Original Features with Financial Ratios for wl = 5

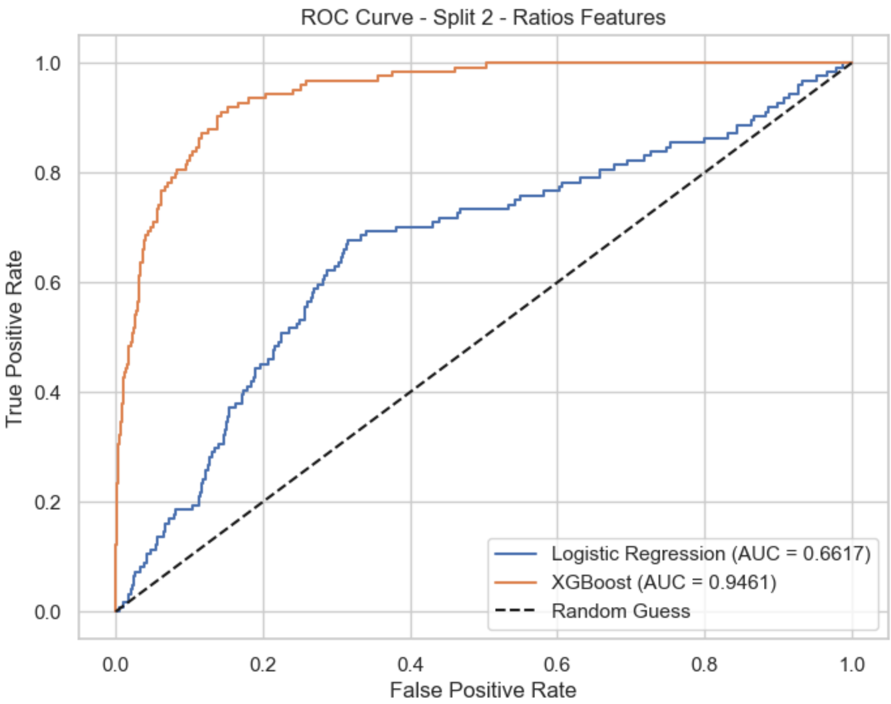


Figure 6. ROC Curve - Split 2 - everything with Macroeconomic Variables for wl = 5

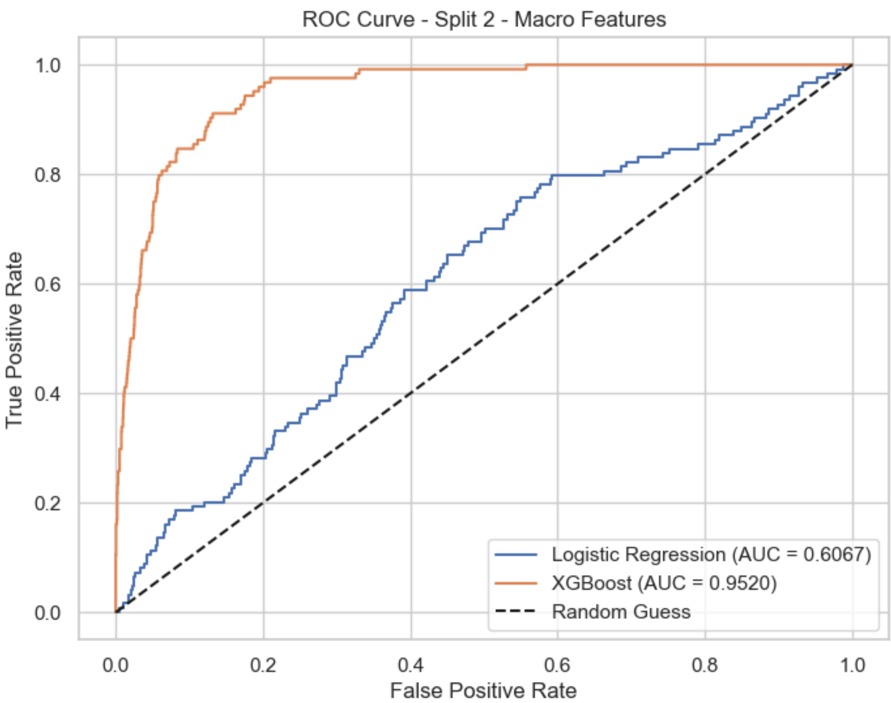


Figure 7. ROC Curve - Split 3 - Original Features for wl = 5

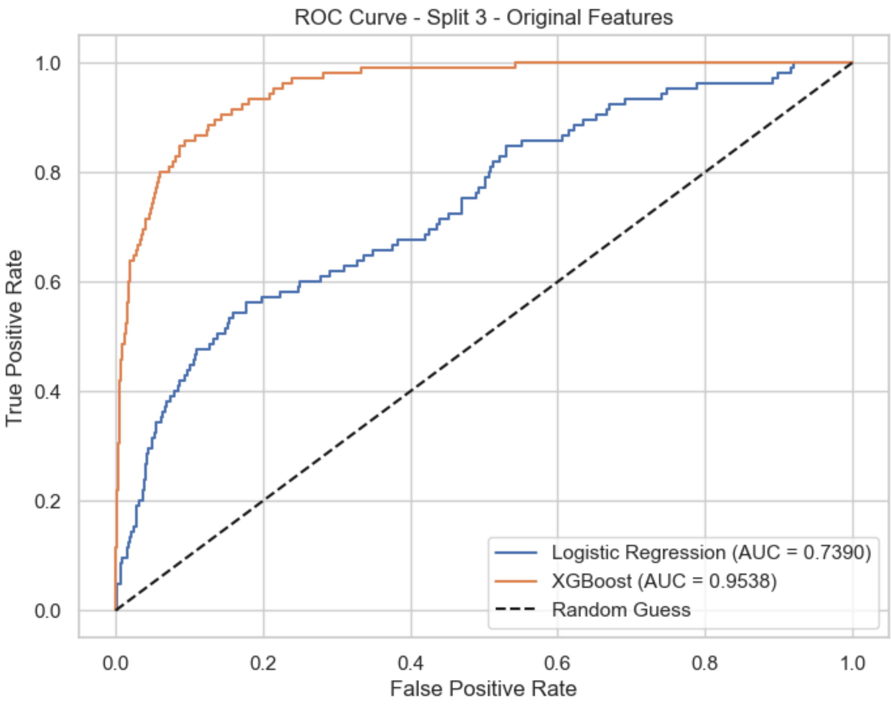


Figure 8. ROC Curve - Split 3 - Original Features with Financial Ratios for wl = 5

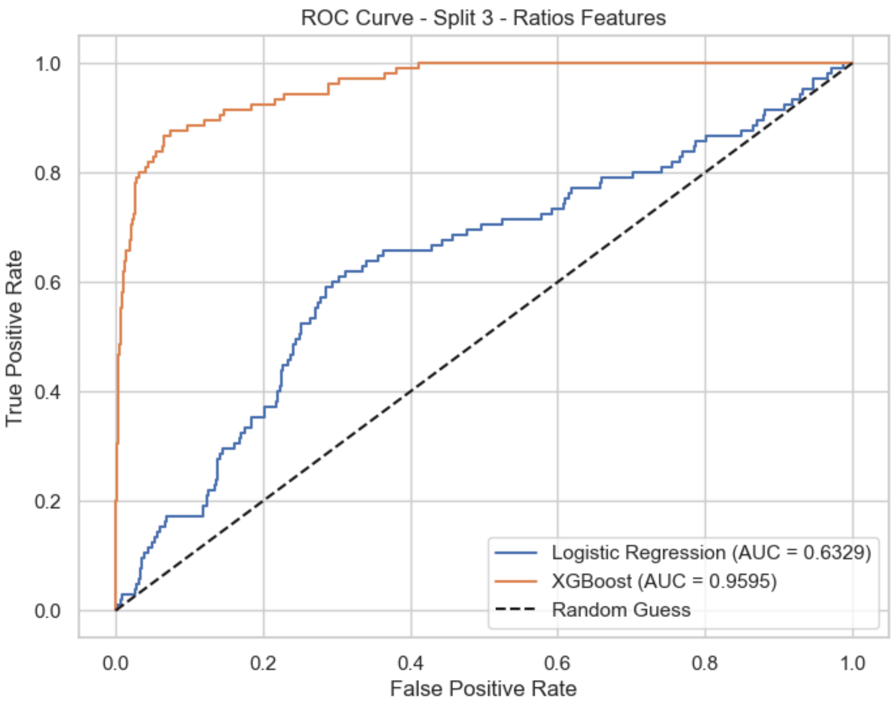


Figure 9. ROC Curve - Split 3 - everything with Macroeconomic Variables for wl = 5

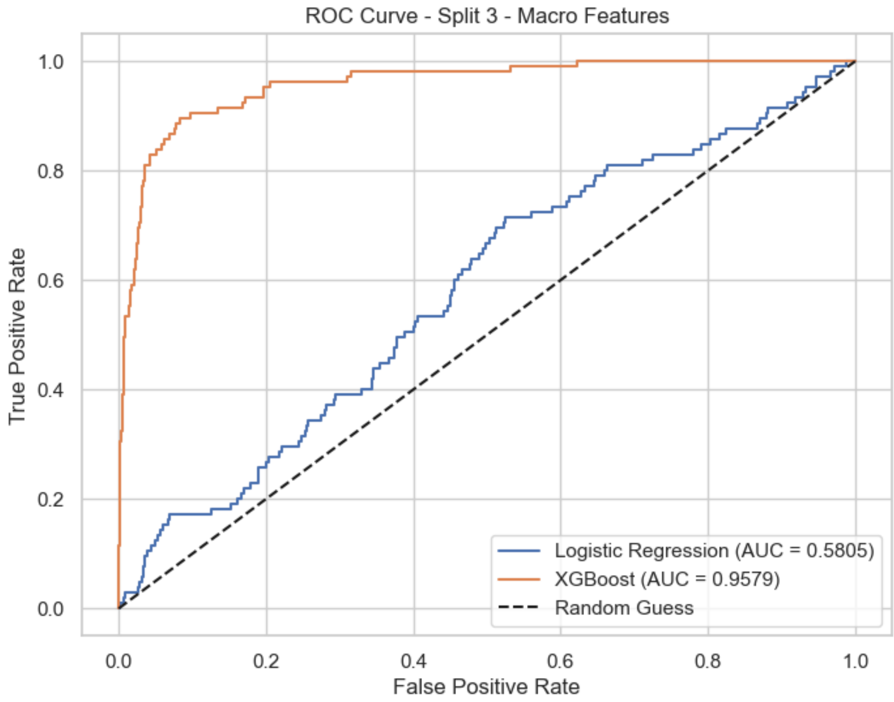


Figure 10. ROC Curve - Split 4 - Original Features for wl = 5

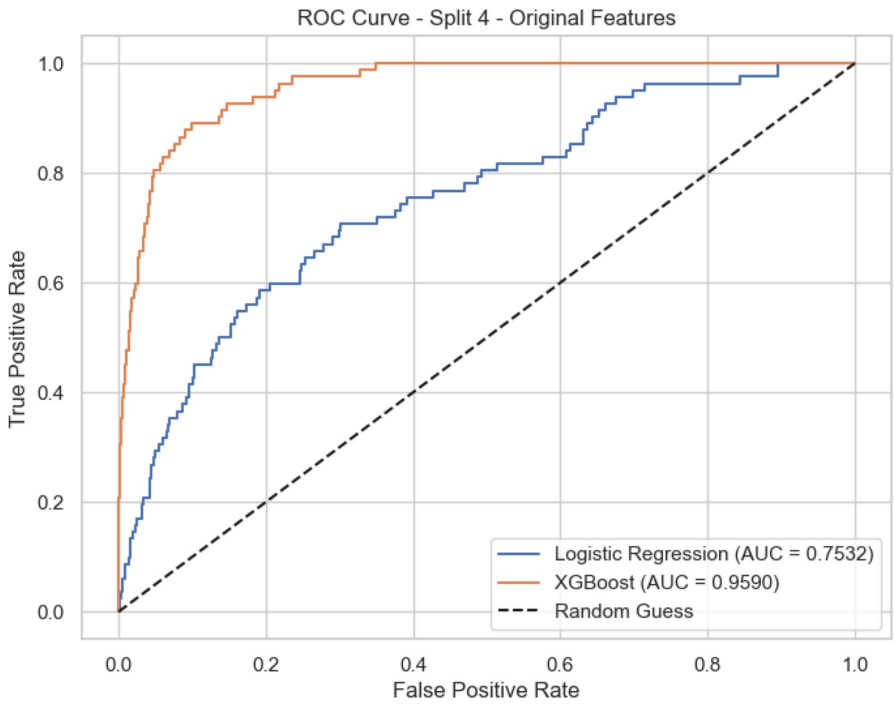


Figure 11. ROC Curve - Split 4 - Original Features with Financial Ratios for wl = 5

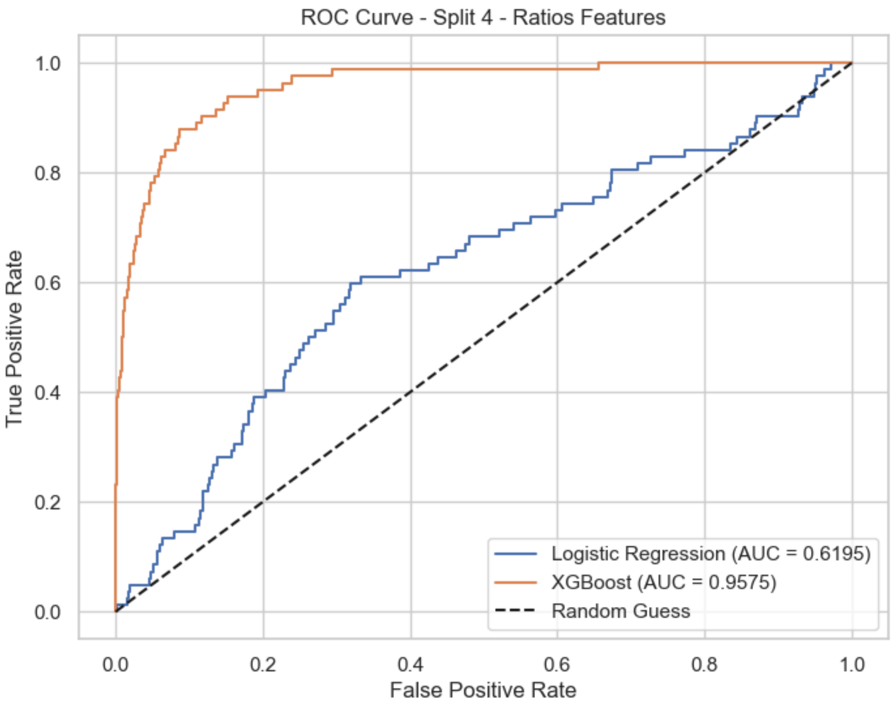


Figure 12. ROC Curve - Split 4 - everything with Macroeconomic Variables for wl = 5

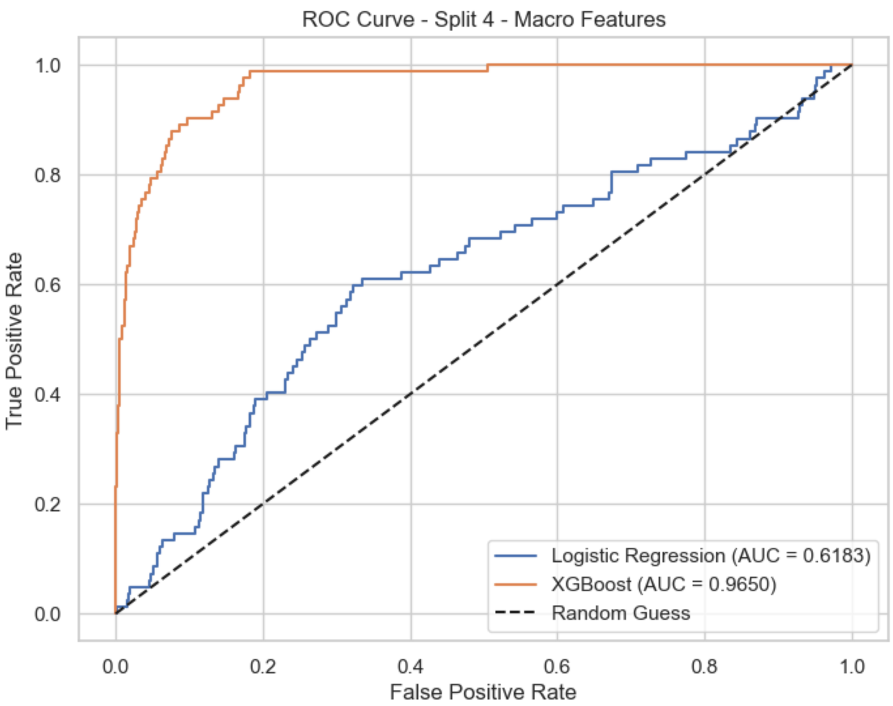


Figure 13. ROC Curve - Split 5 - Original Features for wl = 5

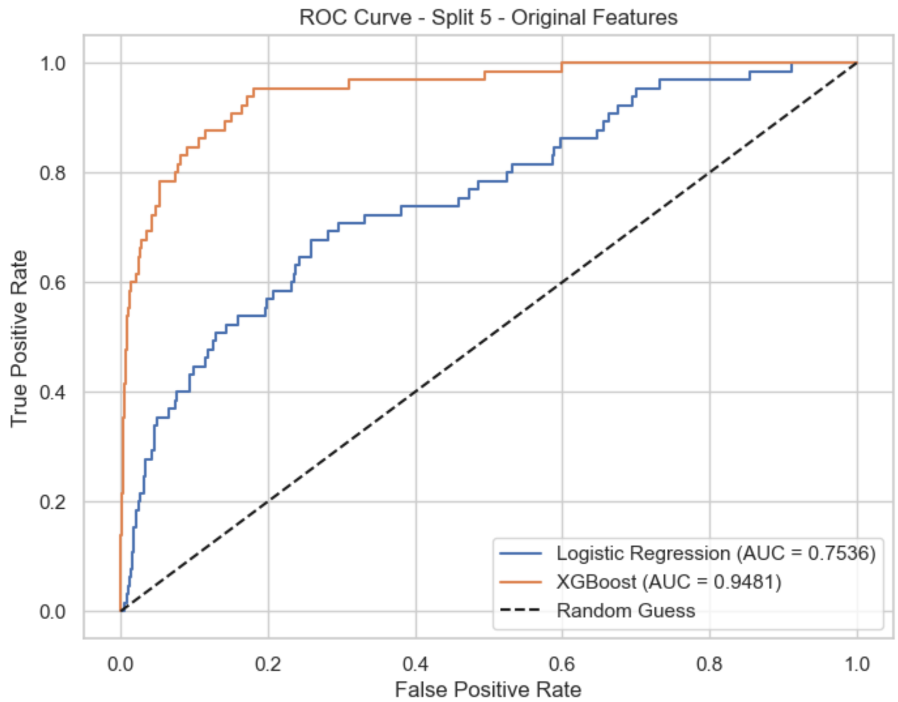


Figure 14. ROC Curve - Split 5 - Original Features with Financial Ratios for wl = 5

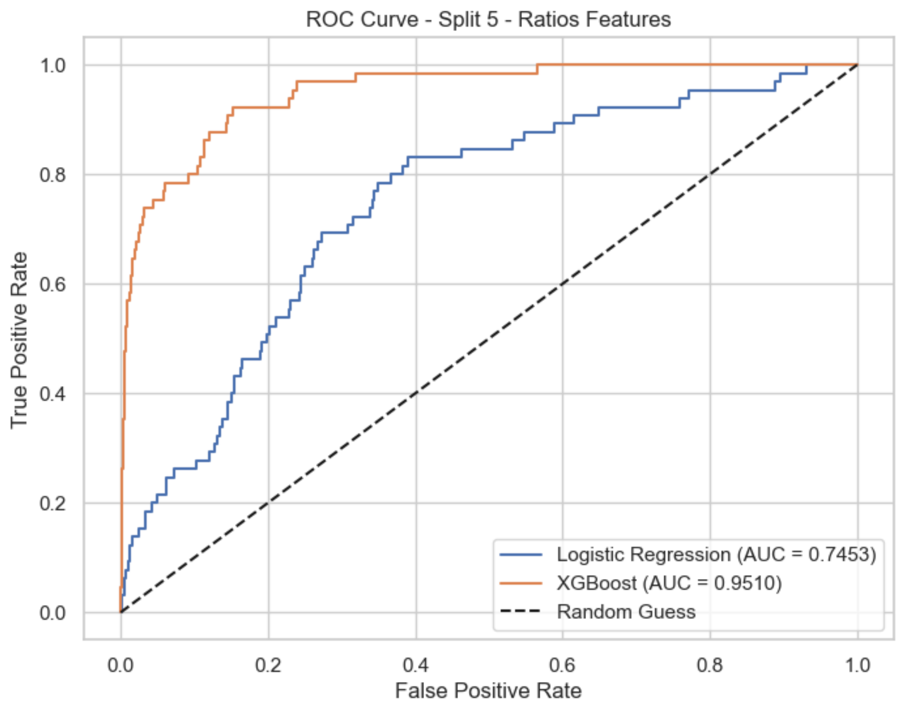


Figure 15. ROC Curve - Split 5 - everything with Macroeconomic Variables for wl = 5

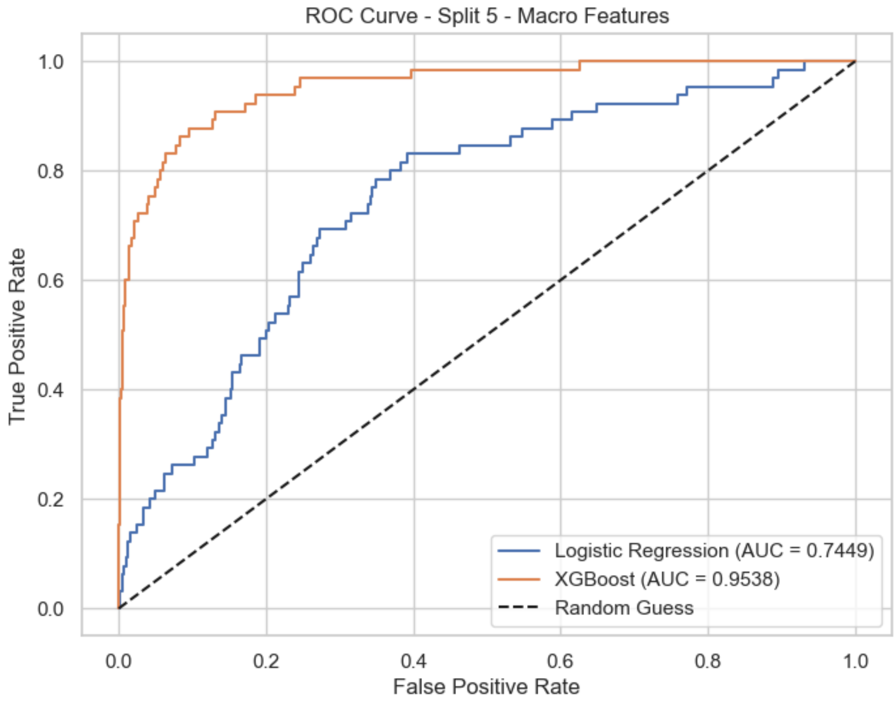


Figure 16. ROC Curve - Split 6 - Original Features for wl = 5

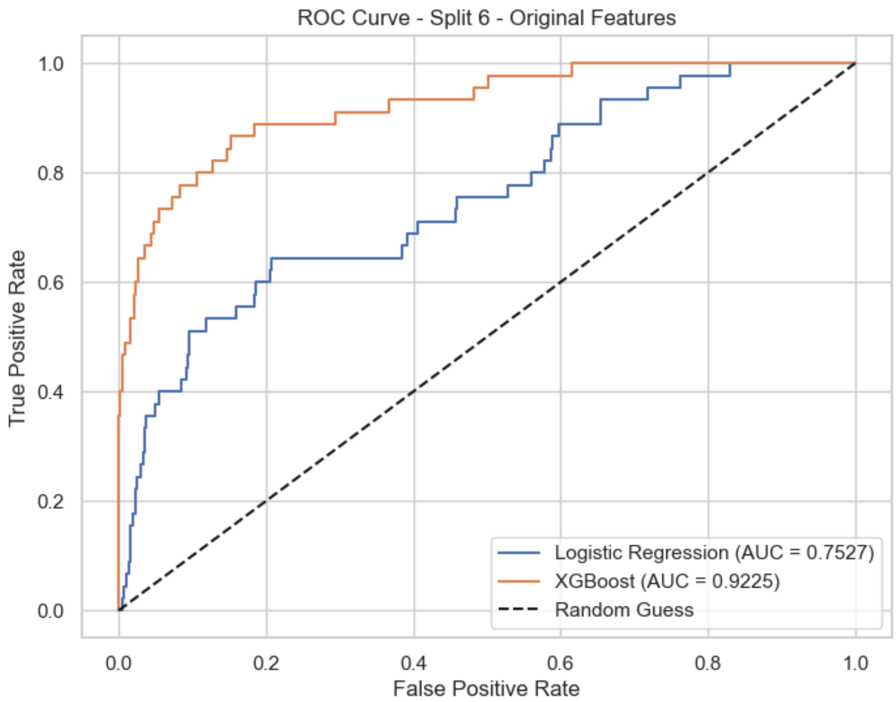


Figure 17. ROC Curve - Split 6 - Original Features with Financial Ratios for wl = 5

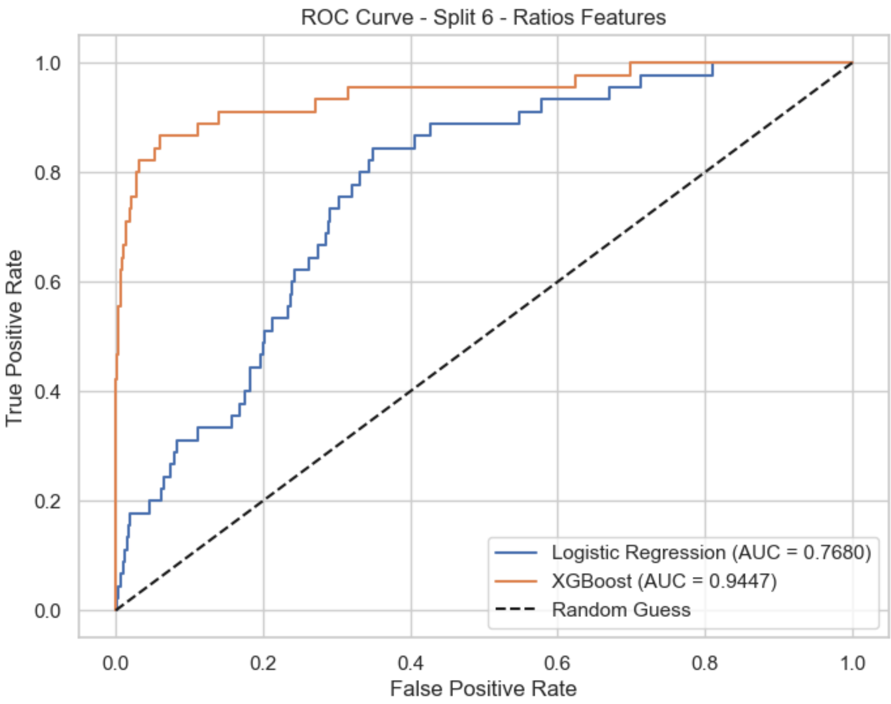


Figure 18. ROC Curve - Split 6 - everything with Macroeconomic Variables for wl = 5

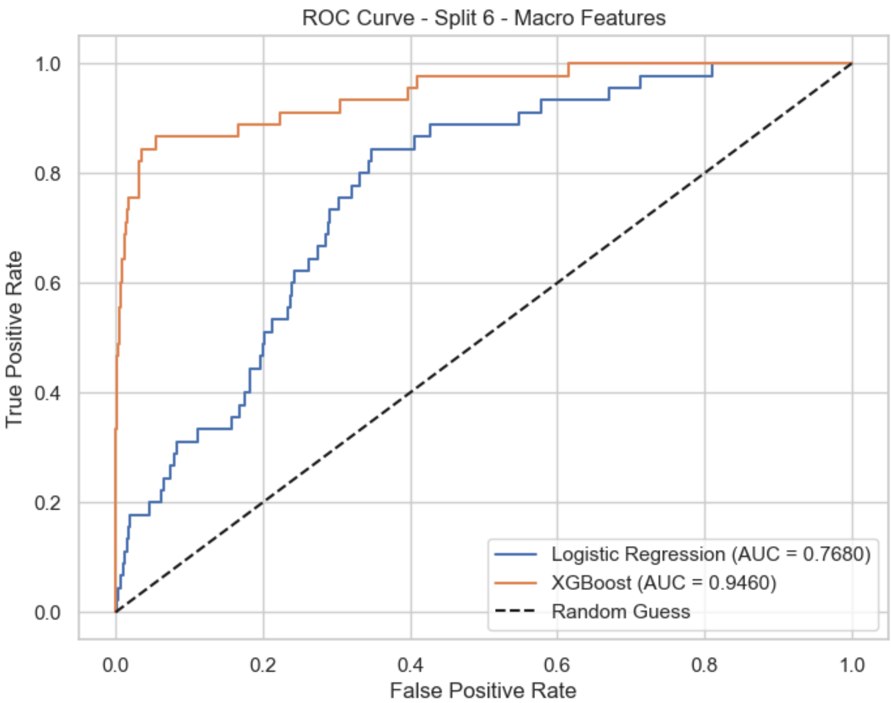


Figure 19. ROC Curve - Split 7 - Original Features for wl = 5

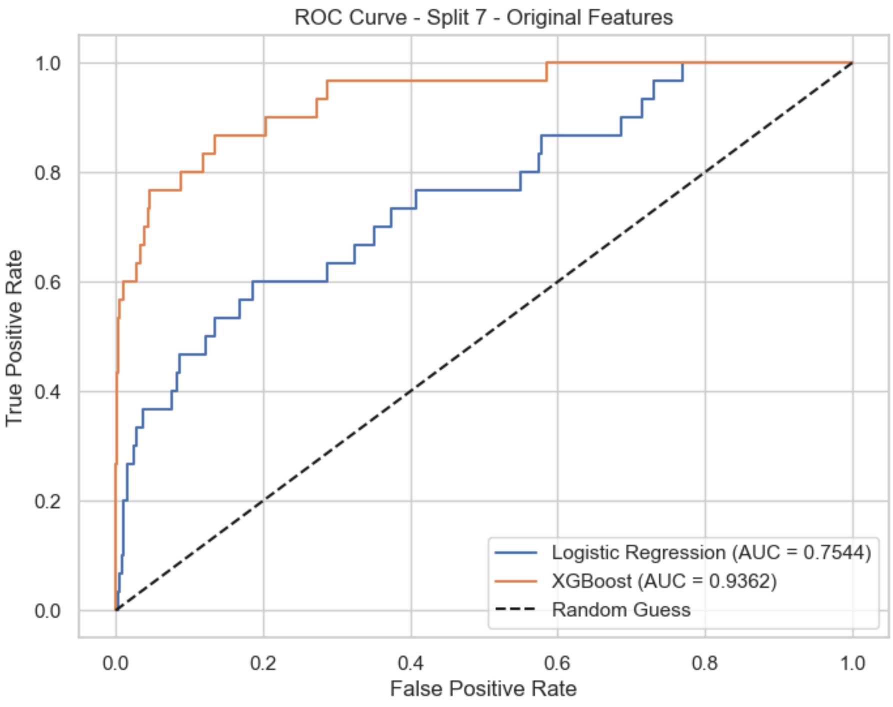


Figure 20. ROC Curve - Split 7 - Original Features with Financial Ratios for wl = 5

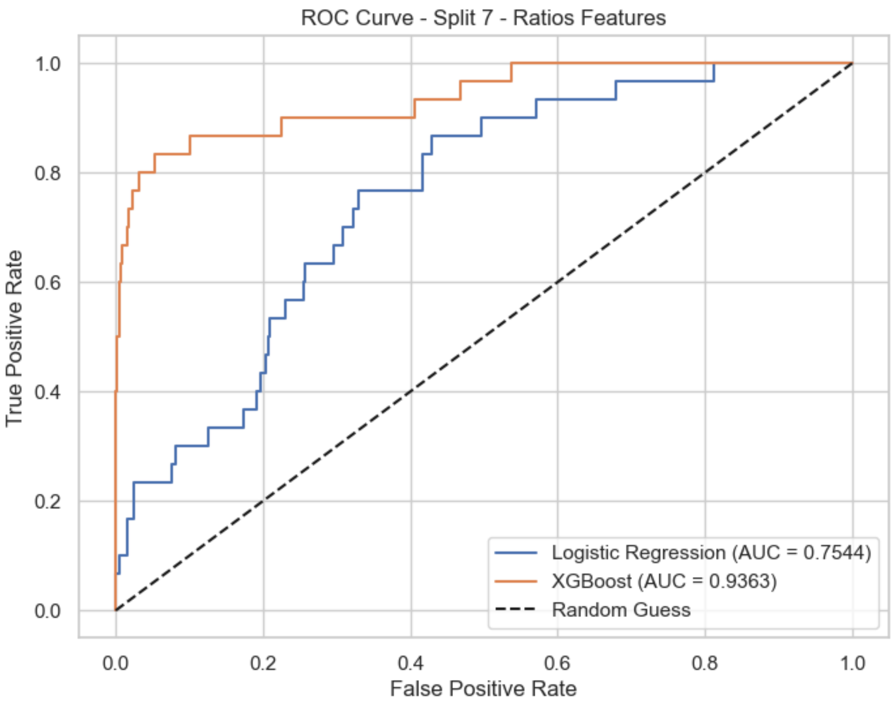
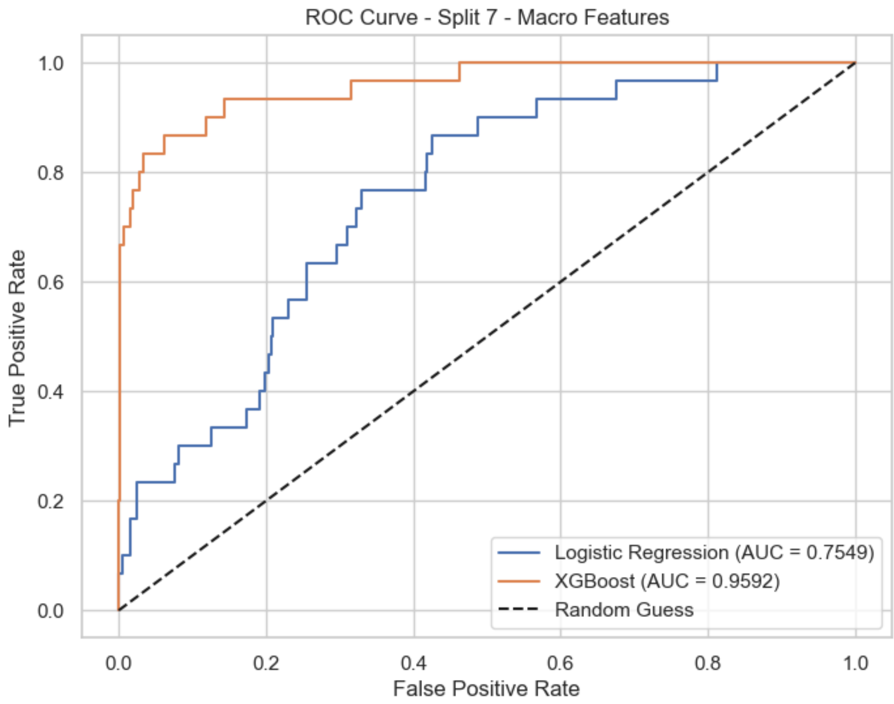


Figure 21. ROC Curve - Split 7 - everything with Macroeconomic Variables for wl = 5



The AUC plots for wl = 9 are provided between *Figure 22* and *Figure 30*.

Figure 22. ROC Curve - Split 1 - Original Features for wl = 9

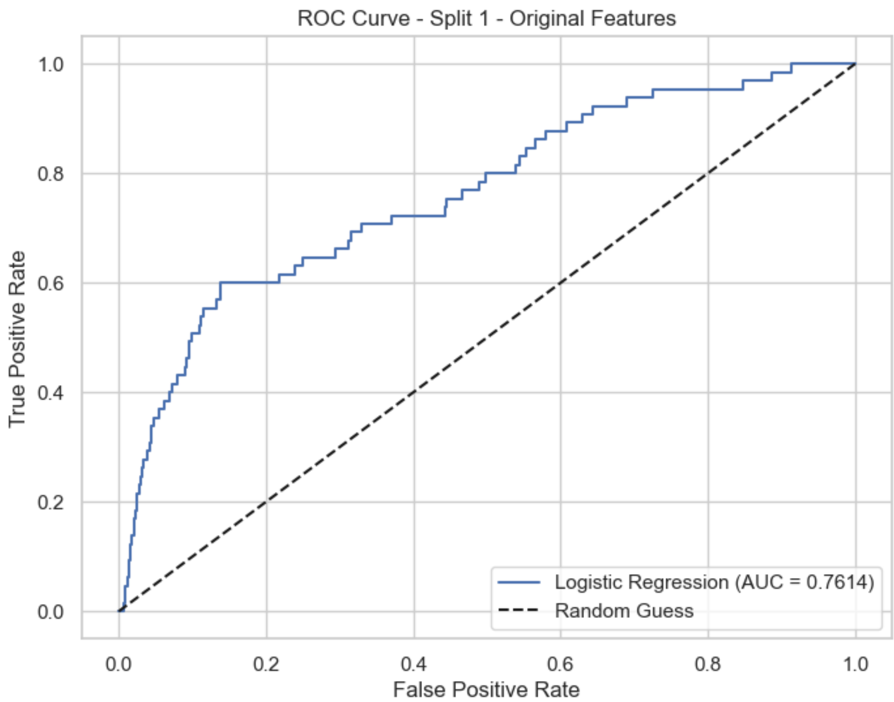


Figure 23. ROC Curve - Split 1 - Original Features with Financial Ratios for wl = 9

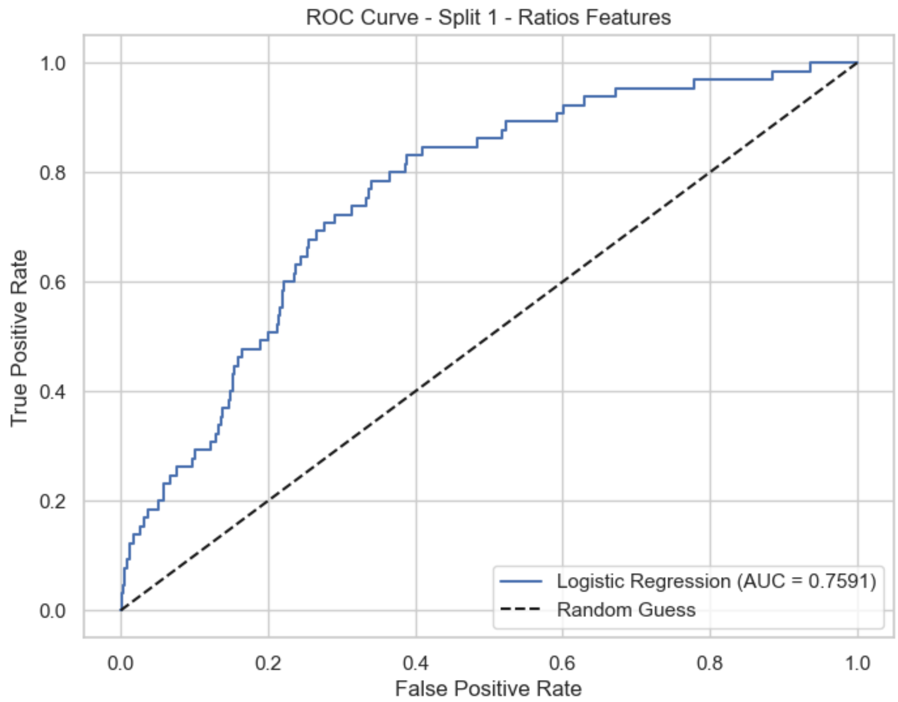


Figure 24. ROC Curve - Split 1 - everything with Macroeconomic Variables for wl = 9

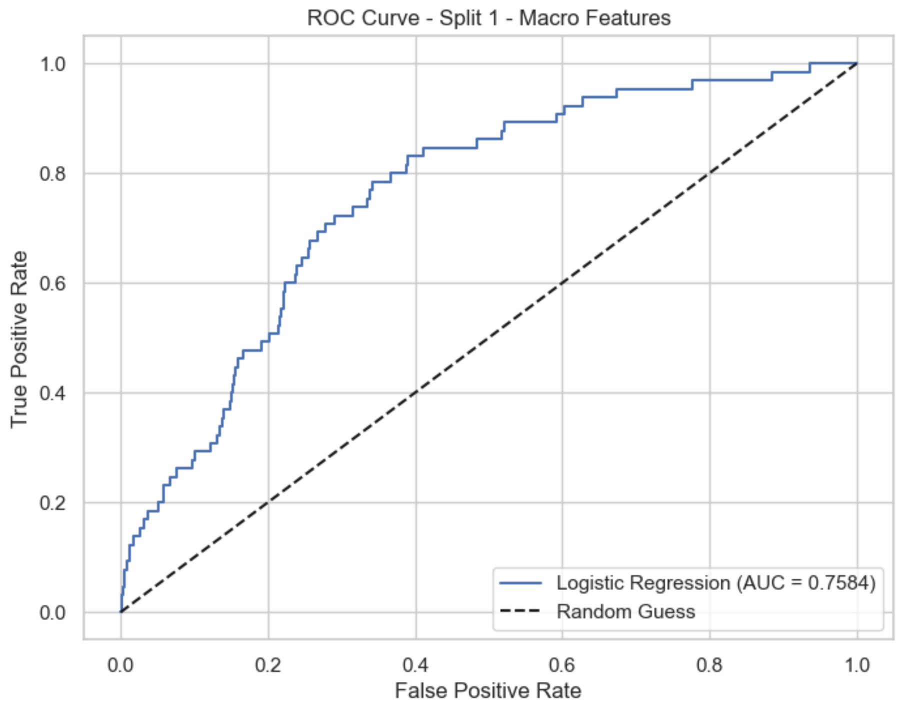


Figure 25. ROC Curve - Split 2 - Original Features for wl = 9

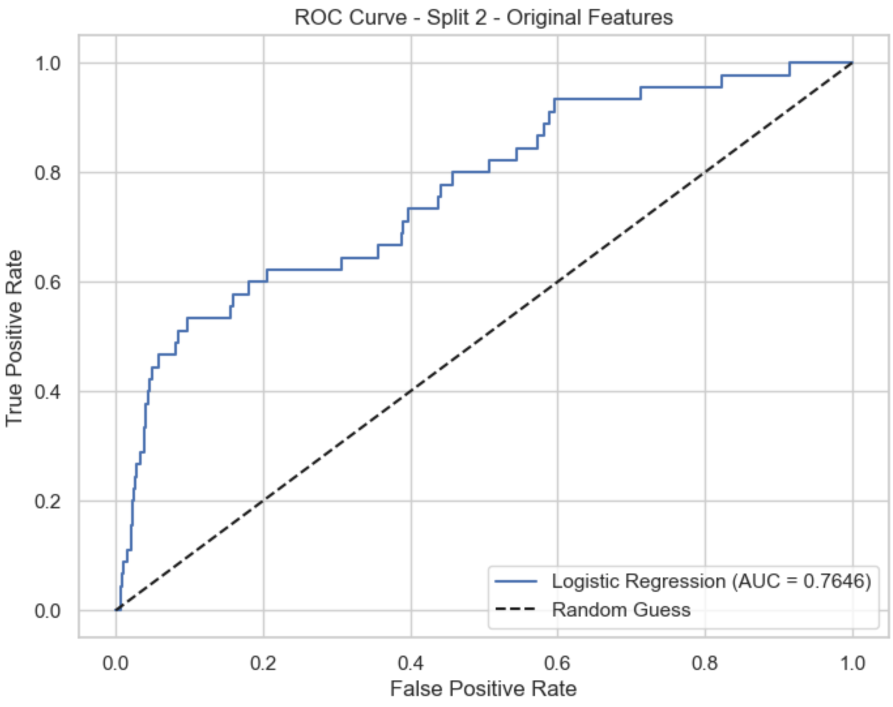


Figure 26. ROC Curve - Split 2 - Original Features with Financial Ratios for wl = 9

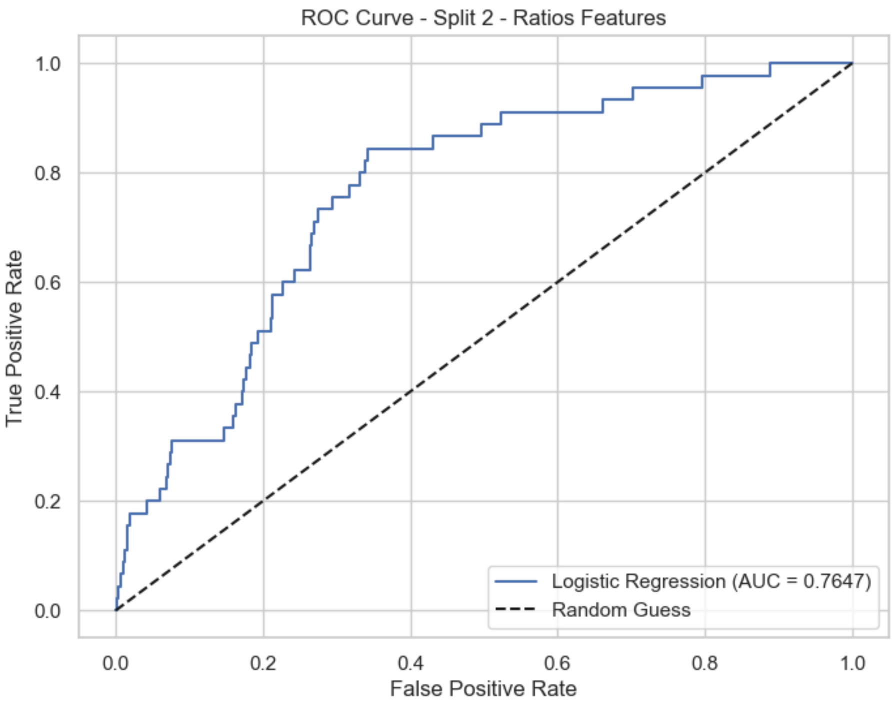


Figure 27. ROC Curve - Split 2 - everything with Macroeconomic Variables for wl = 9

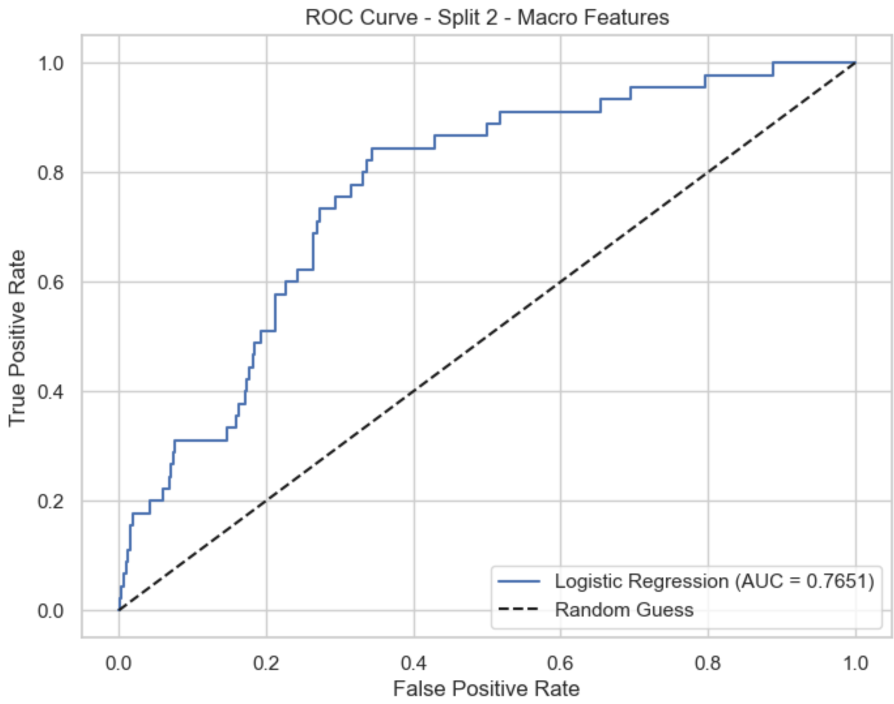


Figure 28. ROC Curve - Split 3 - Original Features for wl = 9

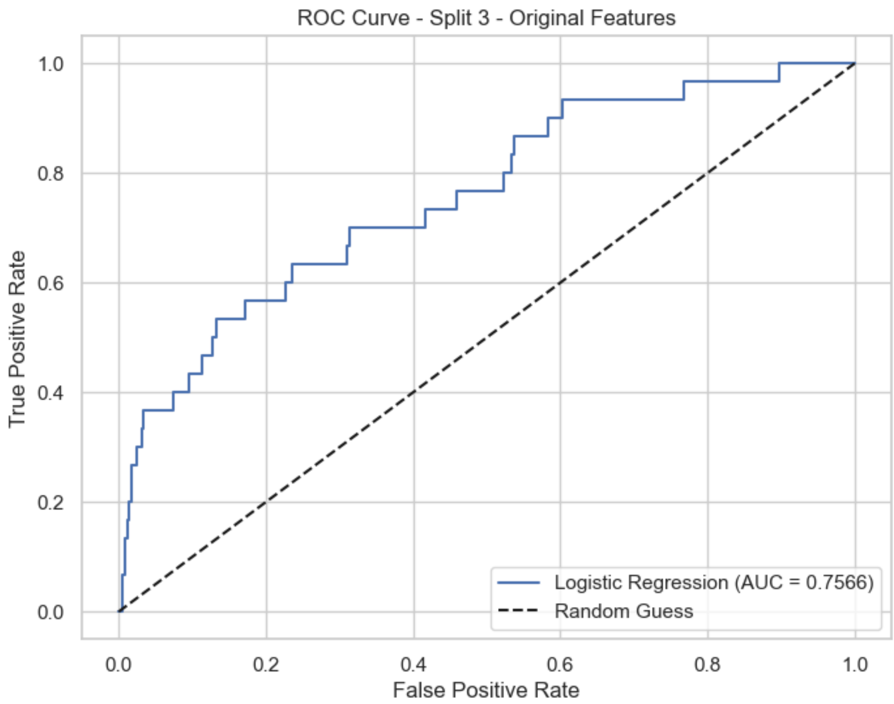


Figure 29. ROC Curve - Split 3 - Original Features with Financial Ratios for wl = 9

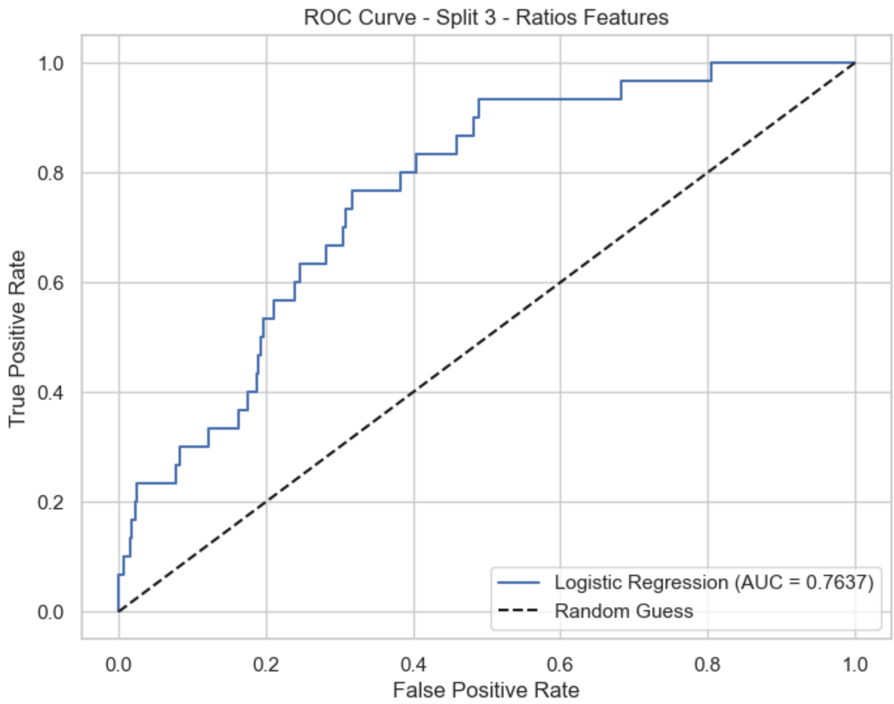


Figure 30. ROC Curve - Split 3 - everything with Macroeconomic Variables for wl = 9