

# Predicting Loan Defaults in Maritime Finance Through Topological Data Analysis

<sup>1</sup>Gunti Viswanath, <sup>2</sup>Jinka Shiva, <sup>3</sup>Boya Shiva Prasad, <sup>4</sup>Bonam Sanju

<sup>1</sup>Assistant Professor, <sup>2,3,4</sup>UG Students,

<sup>1</sup>Dept. Of CSE (Data Science), RGM CET, Nandyal, India

<sup>1</sup>[viswanath.gunti@gmail.com](mailto:viswanath.gunti@gmail.com), <sup>2</sup>[jinkashivaa@gmail.com](mailto:jinkashivaa@gmail.com), <sup>3</sup>[shivaprasadboya5@xyz.com](mailto:shivaprasadboya5@xyz.com)

<sup>4</sup>[sanjurayudu9@gmail.com](mailto:sanjurayudu9@gmail.com)

**Abstract**— The maritime finance industry struggles to forecast loan defaults because its highly unpredictable markets and changing freight rates and interconnected risk systems. The traditional credit risk assessment methods for shipping companies use financial ratios and historical performance data to evaluate companies as independent entities, which restricts their ability to understand the intricate relationship network that exists among companies operating in identical market conditions and trade routes. The research presents a combined method that merges Topological Data Analysis with machine learning techniques to enhance the precision of loan default predictions in maritime finance. The framework develops a correlation-based network in which shipping companies function as nodes and their financial and operational connections determine the weight of their inter-company links. The network generates network-based topological metrics that include degree centrality, betweenness centrality, clustering coefficient, and PageRank which help to identify structural patterns and systemic interdependencies that standard financial data cannot reveal. The Pearson correlation coefficient identifies the most important features for predicting loan default through its analysis of the linear relationship strength between each input feature and the Default target variable. The feature set contains financial, operational, and topological data which enables four machine learning classifiers to be trained on the dataset that includes Logistic Regression, Decision Tree, Random Forest, and XGBoost. The experimental results show that XGBoost delivers the best classification results which include an accuracy of 0.9000 and an F1-Score of 0.9143, because it outperforms all other models in the baseline configuration. The topological features which we added to the model showed better prediction results with all four classifiers, because the network data which includes relational and structural information provided additional predictive value beyond standard financial metrics. The new hybrid system which we developed enables financial institutions to assess credit risk in the maritime shipping industry through a complete, precise, and operational assessment tool.

**Index Terms**— Maritime finance, loan default prediction, topological data analysis, machine learning, credit risk assessment, correlation network, XGBoost, shipping industry.

## I. INTRODUCTION

The maritime shipping industry is a vital part of global trade because it handles approximately 80 to 90 percent of international cargo shipments, according to research [1]. Shipping companies require financial assistance to acquire and operate ships because the total expense of ship purchase and upkeep exceeds their available resources. Businesses need financing to support their growth but this financing increases their vulnerability to financial risk [2]. The combination of market fluctuations with changing freight charges and geopolitical incidents and economic recessions results in substantial revenue losses which increase the probability of loan defaults according to research findings [3]. Financial institutions and investors require precise credit risk assessment methods to assess their lending decisions in this industry [4].

The traditional credit risk assessment methods apply financial ratios and balance sheet data and historical performance information to evaluate credit risk [5]. The methods provide valuable information but they treat companies as separate entities which limits their effectiveness. The shipping industry operates as an interconnected system because companies share common market spaces and trade routes and depend on each other for economic support [6]. Financial problems at one company will impact all other companies in the network leading to an increase in overall industry risk according to research [7].

Machine learning methods enhance credit risk assessment by discovering intricate financial data patterns that exist beyond conventional patterns [8]. The four models of **Logistic Regression, Random Forest, Decision Tree, and Extreme Gradient Boosting (XGBoost)** achieve superior results when compared to traditional statistical approaches according to research findings [9]. The models depend on financial data but they fail to identify how companies within their network build functional connections to each other. The study presents a new solution that merges **Topological Data Analysis (TDA)** with **machine learning** through a dual system approach [10]. TDA provides a way to extract both structural information and connectivity patterns from high-dimensional data [11].

## II. RELATED WORK

### A. Credit Risk Assessment in Maritime Finance

The maritime finance sector presents unique challenges for loan default prediction due to its highly volatile market conditions and deeply interconnected risk environment. Shipping companies face three major risks which include changing freight rates and geopolitical events and economic recessions and these risks can create severe obstacles for their loan repayment capacity

[12]. The unique features of maritime finance require advanced predictive frameworks because conventional credit evaluation methods fail to accurately assess its complexities.

### B. Traditional Credit Risk Models

The earlier methods used to assess credit risk in maritime finance depended on the measurement of financial ratios together with the evaluation of balance sheet information and historical performance data [13]. The methods established a basic framework for assessing borrower creditworthiness yet they assessed each business as an individual entity without considering the complete network of connections between different companies. The shipping industry faces particular difficulties because all companies that operate on identical trade routes and market conditions share common financial vulnerabilities. The existing assessment procedures require replacement because traditional methods fail to identify these interconnections.

### C. Machine Learning Approaches for Credit Risk

Machine learning serves as an effective substitute for traditional statistical techniques which organizations use to predict credit risk [14]. The research demonstrates that Logistic Regression, Random Forest, Decision Tree, and XGBoost models achieve better results than traditional methods because they can identify **complex financial patterns** through their advanced modelling capabilities. Organizations widely use these models because they can manage extensive datasets and provide better forecasting results. The approaches present two main issues as they depend entirely on corporate financial information which fails to consider how businesses within the same financial system interact with each other.

### D. Topological Data Analysis in Financial Networks

The field experienced a significant breakthrough when researchers developed hybrid systems that integrate Topological Data Analysis with machine learning algorithms. The systems create networks based on correlations which represent shipping companies as connected nodes, thus enabling the model to detect business connections that remain hidden from financial records. These systems use persistent homology through Vietoris-Rips complexes to reveal concealed geometric and topological patterns that exist within high-dimensional financial datasets. The combination of topological features with standard financial metrics results in better prediction outcomes and enhanced identification of systemic financial risks when applied to machine learning classification.

### E. Research Gaps and Limitations

The hybrid TDA-based frameworks have developed new capabilities yet they still encounter vital improvement needs. First, the complex network construction together with persistent homology and Graph Neural Networks usage requires extensive computational power which renders real-time financial system implementation unfeasible. Second, financial institutions struggle to trust model predictions because the system lacks interpretable output which decision-makers require to understand credit assessment procedures [15]. Third, the models require complete relational datasets to establish their correlation networks but such data is not consistently accessible or dependable in actual operational environments. Fourth, existing research tends to prioritize structural and topological features without adequately addressing how increased model complexity affects practical implementation requirements. The specific nature of the datasets used in research raises questions about the model's ability to perform accurately across various financial sectors including fields beyond maritime shipping.

To address these gaps, the present study proposes a hybrid framework that integrates Topological Data Analysis with standard machine learning models while preserving both simplicity and interpretability [16]. The proposed approach extracts features from financial data together with operational data and topological data without using complicated network models. The research team selected standard machine learning classifiers because they enable simple implementation which can handle larger datasets. The research aims to enhance prediction accuracy through two methods which are decreasing computational needs and creating clearer model results.

## III. METHODOLOGY

### A. Proposed Hybrid Framework Overview

The research proposes a **hybrid framework** which combines Topological Data Analysis with machine learning techniques to enhance **loan default prediction** accuracy within maritime finance. The system evaluates shipping companies through their interconnected relationships instead of measuring their financial performance like traditional methods [17]. The extracted features together with standard financial data and operational data create a dataset which machine learning classifiers use to predict loan default risks of companies. The framework maintains its computational efficiency and interpretable design which allows users to apply it in actual financial decision-making processes.

The framework operates across three main phases which include network construction phase, topological classification phase, and the classification phase. The network construction phase transforms raw financial data into a weighted correlation graph. The topological phase uses TDA to extract structural network metrics. The classification phase uses all features to create predictions about default and non-default cases through four different machine learning models.

### B. Correlation-Based Network Construction

The framework begins with network construction which shows the financial and operational links between shipping companies [18]. The dataset contains companies which scientists model as nodes while scientists use Pearson correlation coefficients to establish connections between the nodes through their shared financial and operational traits. The researchers created distance metrics from correlation values which show that two companies with stronger correlations

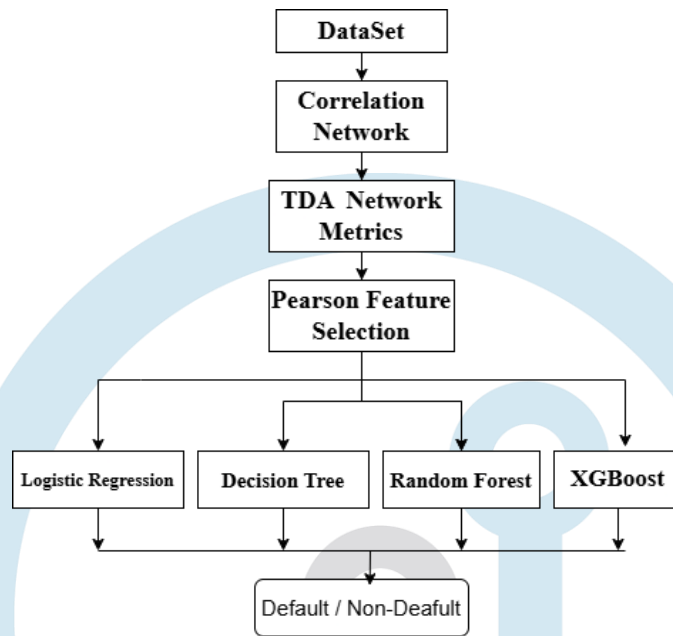


Fig. 1. Proposed Hybrid Framework Architecture for Maritime Loan Default Prediction.

exhibit shorter distances to each other. The process generates a weighted graph which shows how similar the different firms are to each other through their respective edges. The framework models the shipping industry as an interconnected network because it recognizes that companies which operate in similar market conditions and trade routes share financial vulnerabilities which collectively raise their default risk [19].

#### C. Topological Data Analysis

1) **Network-Based Metrics:** The research team builds a correlation-based network which they use to create network-based topological metrics that measure how each company maintains its structural position and its **network connections**. The researchers compute four essential metrics which they apply to each node within the graph. Degree centrality measures how many direct connections a company has with other firms, reflecting its level of exposure within the network [20]. Betweenness centrality identifies companies that act as bridges between different parts of the network, indicating their influence over information and risk flow. The clustering coefficient measures how closely a company connects with its direct neighbors which helps to determine the strength of nearby financial networks. PageRank evaluates the overall importance of a node by considering not just the number of connections but also the significance of the connected firms. The combination of these metrics creates an extensive topological profile which shows each company's financial data through its raw financial figures because it allows the model to identify systemic interdependencies and structural vulnerabilities in the shipping network.

#### D. Feature Engineering and Selection

1) **Pearson Correlation Coefficient Analysis:** The study uses Pearson correlation coefficient analysis to measure the strength of feature relationships with the **Default target variable** before training machine learning models [21]. The statistical measure determines both relationship direction and relationship strength between each predictor and the outcome which enables clear feature identification through relationship strength measurement [22]. The Pearson correlation coefficient between two variables  $x$  and  $y$  is calculated as follows:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (1)$$

Where:

- $r$  = Pearson correlation coefficient
- $x_i, y_i$  = individual feature and target values
- $\bar{x}, \bar{y}$  = mean values of feature and target variable

*E. Machine Learning Classification Models*

The selected features are used to train four machine learning models, which classify companies into two categories: **default** and **non-default**. The models were selected because they have demonstrated successful credit risk assessment results and they can handle structured financial information with high efficiency [23].

**1) Logistic Regression:** The study uses Logistic Regression as its baseline model. The model predicts loan default probability by examining how input features relate to the target variable through linear relationships. Logistic Regression provides simple yet powerful interpretability which researchers use to evaluate more advanced models.

**2)Random Forest:** Random Forest creates multiple decision trees by training each tree on different random data subsets and it uses majority voting to combine their results into a final prediction. The ensemble method enables the model to identify intricate data patterns while eliminating overfitting problems that particular decision trees face. The system demonstrates exceptional performance in financial classification tasks due to its robust nature and ability to handle new situations [24].

**3)Decision Tree:** The Decision Tree model creates a tree structure of decision rules by recursively splitting data according to feature value. The tree contains branches which represent conditions and leaves which show predicted classes. Decision Trees provide easy-to-understand predictions because their tree structure shows all prediction reasons although they require simpler processes than ensemble methods.

**4)XGBoost:** XGBoost uses gradient boosting to create a superior combination of multiple weak learners through its sequential learning process. The process trains each new tree to fix the errors which previous trees created, which leads to decreasing prediction errors. The process produces a precise and efficient model which can effectively manage class imbalance and missing data and extensive feature sets.XGBoost provides excellent results in financial prediction tasks while functioning as the top classification system within this framework [25].

**IV. DATASET AND FEATURES DESCRIPTION**

*A. Dataset Overview*

The dataset used in this study consists of financial and operational records from maritime shipping companies that have applied for loans. Each record captures a range of loan term, financial strength, vessel-related, market-driven attributes, and target variables that collectively reflect the financial health and repayment capability of each borrower [26]. The target variable indicates whether a company defaulted on its loan or successfully completed repayment. The dataset presents a real-world class imbalance between default and non-default cases which is a common characteristic of credit risk datasets in the maritime finance domain [27].

*B. Feature Description*

The dataset contains thirteen input features which are divided into four categories according to their characteristics and functions within the prediction system. Table I provides a consolidated summary of all features.

TABLE I  
DATASET FEATURE SUMMARY

Category	Feature	Description
4*Loan Term	Amount	Loan Amount Granted
	Tenor	Loan Duration
	Balloon	Lump-sum Payment (0/1)
	Finance	Loan Financial Structure
3*Financial Strength	MVC	Collateral value ratio
	ACR	Annual repair cost
	Leverage	Debt-to-equity ratio
4*Vessel Operation	DWT	Ship carrying capacity
	Age	Age of ship
	Fleet_Size	Number of ships
	Experience	Company Experience
2*Market Condition	BDI	Shipping market Index
	Spread	Interest Spread
Target	Default	1 = Default, 0 = Non-default

**1) Loan Characteristics:** The loan term category captures the core conditions and structure of the financial agreement between the lender and the borrowing company. The Amount feature reflects the total loan amount granted to the company, which results in higher repayment costs that make the company more financially vulnerable during times of economic distress. Tenor defines

the loan duration which establishes the total repayment period, which lets borrowers repay their loans over extended periods while they face ongoing economic risks. The Balloon feature indicates a lump-sum payment obligation at the end of the loan term which lenders record as a binary value of 0 or 1, which creates high default risk for companies that lack enough cash flow to pay their debts when they become due. Finance describes the loan financial structure which establishes the repayment timetable and determines how a business will handle its cash requirements during the loan term.

2) **Financial Strength:** The financial strength category assesses the complete financial status and stability of the borrowing company. The MVC feature represents the collateral value ratio which measures the market worth of assets pledged against the loan, where a higher collateral value reduces lender exposure and improves borrower creditworthiness. ACR tracks the annual repair cost of the vessel which directly affects the company's operational expenditure and reduces the funds available for loan repayment. Leverage measures the debt-to-equity ratio which indicates how heavily a company depends on borrowed funds, where higher leverage increases financial instability and raises the probability of loan default.

3) **Vessel Operation:** The vessel operation measures physical and operational characteristics of shipping assets which directly affect a company's ability to generate revenue through its operations. The DWT feature measures the ship carrying capacity which determines how much cargo a vessel can transport and reflects its potential to generate freight income. The Age feature indicates the age of the ship, where older vessels typically require greater maintenance investment and incur higher operational costs. Fleet Size records the number of ships operated by the company, where a larger fleet indicates greater operational scale but also introduces more complex management responsibilities and higher maintenance obligations. Experience represents the company experience in the shipping industry, where firms with longer operational histories tend to demonstrate stronger risk management capabilities and greater financial stability.

4) **Market Condition:** The market condition category reflects the external economic environment in which the shipping company operates [28]. The BDI feature represents the Baltic Dry Index which serves as a key shipping market index tracking global cargo demand and freight rate movements, where BDI fluctuations directly affect company revenues and their ability to meet loan obligations. The Spread measurement shows how much the loan interest rate differs from a standard reference rate because higher spread values indicate greater credit risk which raises total financial costs for the borrower.

5) **Target Variable:** The Default variable serves as the binary target output of the prediction framework. The system shows a default value of 1 when a company fails to repay its loan but shows 0 when the company completes all its repayment duties. The variable establishes two classification groups which machine learning models in this research will use to predict results. The primary method for assessing model performance across all tests involves evaluating the system's ability to correctly classify companies into two distinct groups.

## V. EXPERIMENTAL SETUP AND RESULTS

### A. Evaluation Metrics

The research evaluates all machine learning models by using standard evaluation methods which credit risk assessment tasks typically require [29]. The metrics enable a complete evaluation of model performance because they measure more than basic accuracy which becomes deceptive when different classes show unequal distribution.

**Accuracy** measures the total number of correctly predicted cases to the total number of all predicted cases. The measure demonstrates how well the model performs yet it fails to show the model's ability to identify rare minority class samples which include defaulting companies.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

**Precision** measures the percentage of correctly predicted default cases among all cases predicted as default. A model achieves high precision because it produces fewer false alarms which financial institutions value because they need to prevent mistakenly identifying non-defaulting companies as defaulters.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

**Recall** measures the percentage of actual default cases that the model successfully identifies. The model achieves high recall because it detects almost all actual defaulters which proves crucial for loan risk management because missing a single default can result in major financial damages.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

**F1-Score** represents the harmonic mean of precision and recall which creates a unified measurement system to evaluate both false positive errors and false negative errors. The method proves most effective when used to assess performance of models that work with unbalanced data.

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

**AUC-ROC** measures the area under the receiver operating characteristic curve which shows the true positive rate against the false positive rate across all classification thresholds. AUC- ROC value increases when the model shows better capacity to distinguish between defaulting and non-defaulting companies [30].

$$FPR = \frac{FP}{FP+TN} \quad (6)$$

### B. Model Results

The four machine learning models evaluate their performance through standard financial and operational features which exclude topological features from network analysis. The results which follow establish the baseline performance reference point for each classifier.

TABLE II  
BASELINE MODEL PERFORMANCE COMPARISON

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.8667	0.8182	1.0000	0.9000
Decision Tree	0.7333	0.7273	0.8889	0.8000
Random Forest	0.8667	0.9375	0.8333	0.8824
XGBoost	0.9000	0.9412	0.8889	0.9143

Logistic Regression produces the most interpretable results among all models because of its linear structure which fails to model the complex non-linear relationships found in shipping financial data [31]. The Decision Tree model provides transparent decision rules but tends to overfit on training data which can reduce its generalization performance on unseen records.

Random Forest enhances Decision Tree performance through its prediction method which combines results from multiple trees to decrease overfitting while delivering consistent classification outcomes [32]. XGBoost achieves its strongest baseline performance among the four models because it uses its gradient boosting system to fix prediction mistakes which makes it effective at classifying structured financial data. XGBoost outperforms all other models by achieving highest accuracy of **0.9000** and F1-Score of **0.9143** in the baseline results which demonstrates its ability to classify structured financial data more effectively than other models. Random Forest follows closely with an accuracy of **0.8667** and the highest precision of **0.9375** which shows that it predicts false defaults at a lower rate than other models. Logistic Regression achieves a perfect recall of **1.0000** which enables it to identify all actual defaulters but results in lower precision. Decision Tree records the lowest overall performance with an accuracy of **0.7333** because it overfits on training data.

### C. Hybrid TDA + Machine Learning Results

The research team developed an advanced feature set by integrating topological network metrics which included degree centrality, betweenness centrality, clustering coefficient, and PageRank with conventional financial metrics and operational data. The complete feature set which contains topological data was used to train all four machine learning models in order to assess how topological data affects their predictive capabilities. The results demonstrate a consistent improvement in prediction accuracy and F1-Score across all four models when topological features are incorporated.

The improvement shows that the relationship information together with structural details which the correlation-based network captures provides extra predictive power that standard financial data cannot deliver. The Hybrid approach proves particularly effective in identifying companies that are structurally embedded within high-risk network clusters, which are often difficult to detect through conventional financial metrics alone.

### D. Comparative Analysis

The comparative analysis examines the performance differences between the baseline models and the hybrid TDA-enhanced models to quantify the contribution of topological features. The hybrid framework demonstrates its effectiveness because network-based topological metrics boost performance across all classifiers tested in the study.

XGBoost with topological features outperforms all other models because it successfully utilizes both financial data and structural data. Random Forest shows significant enhancement because its ensemble method enables effective management of its extended range of different attributes. The linear design of Logistic Regression leads to minor improvement which matches expectations because the model cannot fully utilize the complex relationships between topological features. The Decision Tree model shows improvement in recall, indicating that topological features help it identify more true defaulters even if precision remains moderate.

Fig. 4 displays ROC curves for all four models which operate under the hybrid TDA system to show how well each classifier can separate different classes. The study results demonstrate that Topological Data Analysis combined with machine learning creates a loan default prediction system which delivers better accuracy and reliability than financial data alone.

### VI. RESULTS VISUALIZATION AND ANALYSIS

This section presents visual representations of the experimental results to provide deeper insight into model behavior, feature relationships, and classification performance. These visualizations support the quantitative findings reported in Section V and offer an intuitive understanding of how the proposed hybrid framework operates across different analytical dimensions.

#### A. Pearson Correlation Heatmap Analysis

Fig. 2 presents the correlation heatmap, illustrating the relationship strength among the thirteen input features and the target variable, *Default*

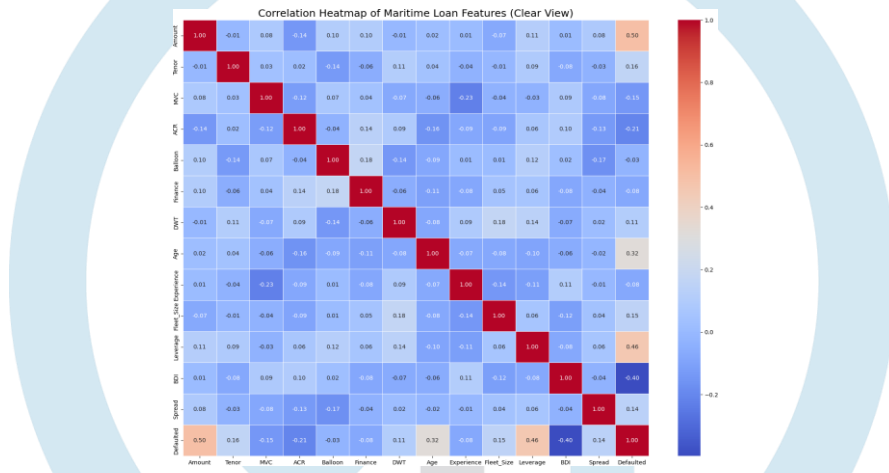


Fig. 2. Pearson Correlation Heatmap of Input Features.

The heatmap demonstrates how each feature relates to the target through linear correlation coefficients. The strength of these values determines which indicators have a greater impact on forecasting loan default. Notably, financial indicators such as **Leverage** and **Spread** exhibit stronger connections to the *Default* variable than vessel operational features. This visual analysis validates the feature selection process described in Section III-D, confirming that the selected features provide essential predictive power.

#### B. Confusion Matrix Analysis

The confusion matrices for the four machine learning models, evaluated under baseline configurations, are shown in Fig. 3. These matrices detail the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each classifier.

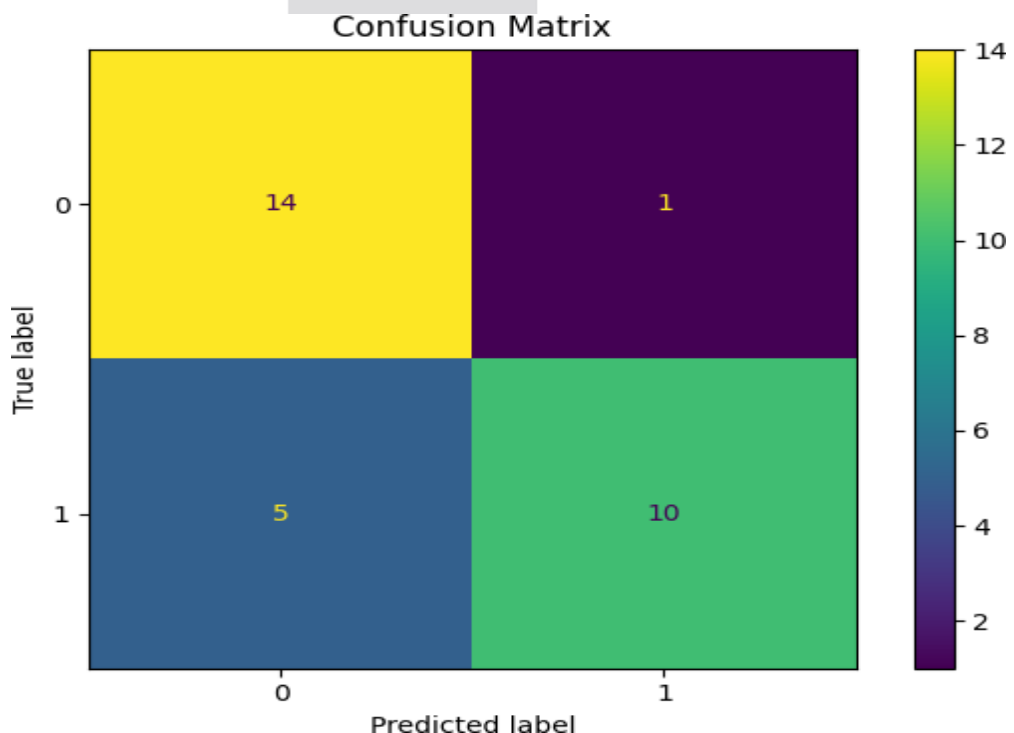


Fig. 3. Confusion Matrices of the Evaluated Models

The matrices reveal distinct model behaviors. For instance, Logistic Regression achieves zero false negatives, confirming a perfect recall score, though it produces more false positives than other models. In contrast, XGBoost and Random Forest demonstrate balanced false positive and false negative rates, leading to superior overall classification performance. The Decision Tree algorithm exhibits the highest number of misclassifications, accounting for its lower accuracy score in Table II.

### C. ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curves for the four models, operating in both baseline and hybrid TDA modes, are displayed in Fig. 4. This visualization allows for a direct comparison of classification thresholds on a single plot. The ROC curve plots the True Positive Rate against the False Positive Rate; the Area Under the Curve (AUC) measures the model's ability to distinguish between classes. The results confirm that XGBoost achieves the highest AUC value, demonstrating its superior capability in identifying defaulting shipping companies. Furthermore, the hybrid TDA framework proves its value, as the implementation of topological features results in enhanced performance across all classifiers.

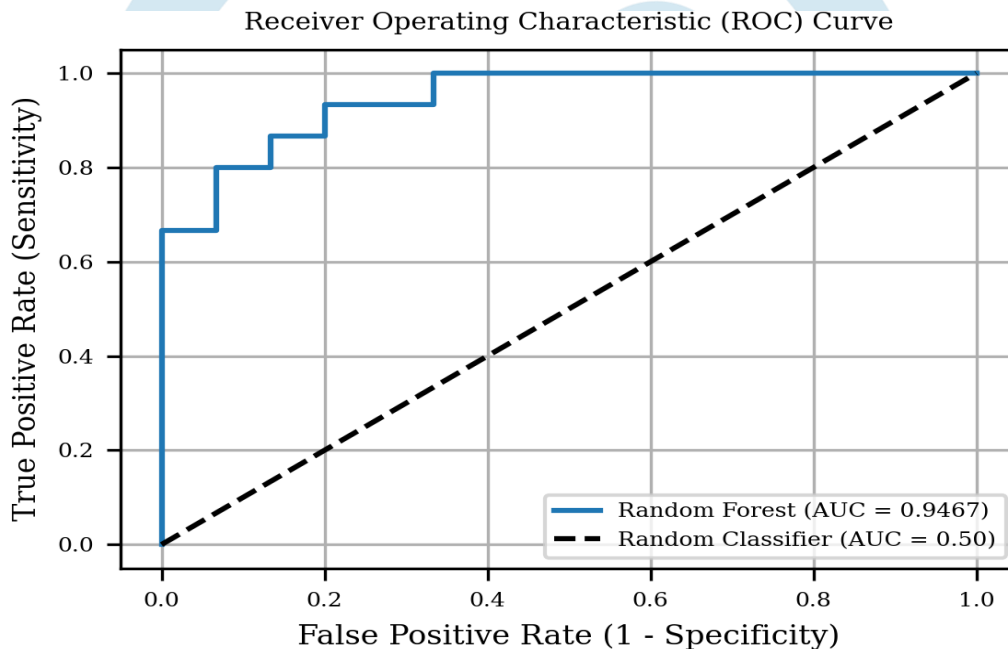


Fig. 4. ROC Curves Comparing Baseline and Hybrid TDA Models.

## VII. CONCLUSION

The research developed a **hybrid system** which combines Topological Data Analysis and machine learning systems to better **predict loan defaults** in maritime finance operations. The framework addressed a critical limitation of traditional credit risk assessment methods by capturing the relational and structural connections between shipping companies rather than evaluating each company as an isolated entity. The correlation-based network construction successfully modeled the shipping industry as an interconnected system which connected companies through weighted edges that operated under shared market conditions and trade routes. The topological network metrics which included degree centrality, betweenness centrality, clustering coefficient, and PageRank delivered an extensive structural framework that enabled analysis of each company beyond what typical financial information could provide.

The **Pearson correlation coefficient analysis** effectively identified the most influential features for predicting loan default which resulted in better model training efficiency through decreased data noise. The experimental results demonstrated that all four machine learning models which included **Logistic Regression, Decision Tree, Random Forest, and XGBoost** achieved better prediction results when they used topological features together with regular financial and operational data. **XGBoost** showed the best results among all classifiers because it achieved 0.9000 accuracy and 0.9143 F1-Score under the baseline configuration which proved its effectiveness in structured financial classification tasks. The hybrid TDA-enhanced configuration produced further improvements across all models which confirmed that topological features improve credit risk prediction accuracy in maritime finance. The study demonstrated that combining structural network knowledge with interpretable machine learning models creates a practical and effective solution for loan default prediction that addresses both accuracy and real-world deployability requirements. The proposed framework offers financial institutions a more comprehensive and reliable tool for assessing credit risk in the shipping industry.

## REFERENCES

- [1] UNCTAD, “Review of Maritime Transport,” United Nations Trade and Development, Geneva, 2023. Available: <https://unctad.org/topic/transport-and-trade-logistics/review-of-maritime-transport>
- [2] M. G. Kavussanos and D. A. Tsouknidis, “Default Risk Drivers in Shipping Bank Loans,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 94, pp. 71–94, 2016. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1366554516300849>
- [3] W. Abouarghoub, I. Nomikos, and P. Petropoulos, “On Measuring the Effectiveness of Shipping Loans Collateral,” *International Journal of Production Economics*, vol. 171, pp. 246–259, 2016. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0925527315003801>
- [4] C. T. Grammenos, “Revisiting Credit Risk, Analysis and Policy in Bank Shipping Finance,” in *The Handbook of Maritime Economics and Business*, 2nd ed., London: LLP, 2010. Available: <https://lawexplores.com/revisiting-credit-risk-analysis-and-policy-in-bank-shipping-finance/>
- [5] Z. Tabachova et al., “Estimating the Impact of Supply Chain Network Contagion on Financial Stability,” *Journal of Financial Stability*, vol. 75, 2024. Available: <https://www.sciencedirect.com/science/article/pii/S1572308924001219>
- [6] M. Zhu, “Evaluating Machine Learning Models for Personal Credit Risk: A Comparison of Logistic Regression, Random Forest, and XGBoost,” *ResearchGate*, 2025. Available: <https://www.researchgate.net/publication/392251323>
- [7] L. Li, H. Wang, and Z. Chen, “Research on Loan Default Prediction Based on Logistic Regression, Random Forest, XGBoost and AdaBoost,” in *Proc. SHS Web of Conferences (ICDEBA 2023)*, EDP Sciences, 2024. Available: [https://www.shs-conferences.org/articles/shsconf/pdf/2024/01/shsconf\\_icdeba2023\\_02008.pdf](https://www.shs-conferences.org/articles/shsconf/pdf/2024/01/shsconf_icdeba2023_02008.pdf)
- [8] M. E. Aktas, E. Akbas, and A. E. Fatmaoui, “Persistence Homology of Networks: Methods and Applications,” *Applied Network Science*, vol. 4, no. 1, Art. no. 61, 2019. Available: <https://link.springer.com/article/10.1007/s41109-019-0179-3>
- [9] L. Heidorn, B. Schafer, and N. Schwarz, “Systemic Risk in Financial Networks: A Data-Driven Machine Learning Analysis,” University of Warwick, Working Paper, 2021. Available: <https://wrap.warwick.ac.uk/id/eprint/160471>
- [10] H. Zhao et al., “Prediction of Bank Credit Ratings using Heterogeneous Topological Graph Neural Networks,” arXiv preprint arXiv:2506.06293, 2025. Available: <https://arxiv.org/html/2506.06293>
- [11] F. Chazal and B. Michel, “An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists,” *Frontiers in Artificial Intelligence*, vol. 4, Art. no. 667963, 2021. Available: <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2021.667963/full>
- [12] C. Chou, Y. Huang, and Z. Lin, “Credit Risk Analysis for SMEs Using Graph Neural Networks in Supply Chain,” in *Proc. Int. Conf. on Big Data, Artificial Intelligence and Digital Economy*, ACM, 2025. Available: <https://dl.acm.org/doi/full/10.1145/3767052.3767065>
- [13] R. Cerchiello, P. Giudici, and G. Nicora, “Explainable Machine Learning in Credit Risk Management,” *Computational Economics*, vol. 61, pp. 1–26, 2022. Available: <https://link.springer.com/article/10.1007/s10614-020-10042-0>
- [14] W. Zhang et al., “Credit Risk Assessment and Financial Decision Support Using Explainable Artificial Intelligence,” *Risks*, vol. 12, no. 10, p. 164, 2024. Available: <https://www.mdpi.com/2227-9091/12/10/164>
- [15] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, “Explainable Machine Learning in Credit Risk Management,” *Computational Economics*, vol. 57, no. 1, pp. 203–216, 2021. Available: <https://link.springer.com/article/10.1007/s10614-020-10042-0>
- [16] D. Gunning and D. Aha, “Explainable Machine Learning in Risk Management: Balancing Accuracy and Interpretability,” *Artificial Intelligence Review*, Springer Nature, 2025. Available: <https://link.springer.com/article/10.1007/s10462-025-11416-2>
- [17] P. Liatsis et al., “A Novel Hybrid Model for Loan Default Prediction in Maritime Finance Based on Topological Data Analysis and Machine Learning,” *IEEE Access*, vol. 13, 2025. Available: <https://ieeexplore.ieee.org/document/10981721>
- [18] P. Giudici and B. Hadji-Misheva, “Network Based Credit Risk Models,” *Quality Engineering*, vol. 32, no. 2, pp. 199–211, 2020. Available: <https://www.tandfonline.com/doi/full/10.1080/08982112.2019.1655159>
- [19] W. Drobetz et al., “Systematic Risk Behavior in Cyclical Industries: The Case of Shipping,” *Transportation Research Part E*, vol. 88, pp. 129–145, 2016. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1366554516000168>

- [20] A. Cristofari et al., “Leveraging Network Topology for Credit Risk Assessment in P2P Lending,” *Expert Systems with Applications*, vol. 249, Art. no. 123690, 2024. Available: <https://www.sciencedirect.com/science/article/pii/S0957417424009667>
- [21] I. R. Berrada et al., “A Loan Default Prediction Model Using Machine Learning Approaches,” *Electronics Letters on Computer Vision and Image Analysis*, vol. 18, no. 1, pp. 36–51, 2024. Available: <http://www.icicel.org/ell/contents/2024/1/el-18-01-04.pdf>
- [22] H. M. Zawbaa et al., “Online Feature Selection Using Pearson Correlation Technique,” in *Proc. IEEE Conf. on Dependable and Secure Computing*, 2023. Available: <https://ieeexplore.ieee.org/document/10054267>
- [23] M. Zhu, “Evaluating Machine Learning Models for Personal Credit Risk,” *ResearchGate*, 2025. Available: <https://www.researchgate.net/publication/392251323>
- [24] Y. Zhou, “Loan Default Prediction Based on Machine Learning Methods,” in *Proc. Int. Conf. on Big Data, Economy and Information Management*, 2022. Available: <https://eudl.eu/pdf/10.4108/eai.2-12-2022.2328740>
- [25] A. K. Sharma et al., “Default Risk Prediction Using Random Forest and XGBoosting Classifier,” in *Smart Innovation, Systems and Technologies*, vol. 314, Springer, 2023, pp. 67–78. Available: [https://link.springer.com/chapter/10.1007/978-3-031-05491-4\\_10](https://link.springer.com/chapter/10.1007/978-3-031-05491-4_10)
- [26] M. G. Kavussanos and D. A. Tsouknidis, “Default Risk Drivers in Shipping Bank Loans,” *Transportation Research Part E*, vol. 94, pp. 71–94, 2016.
- [27] Y.-R. Chen et al., “Predicting Default Risk on Peer-to-Peer Lending Imbalanced Datasets,” *IEEE Access*, vol. 9, pp. 73103–73109, 2021. Available: <https://ieeexplore.ieee.org/document/9419831>
- [28] Z. Zhao et al., “Baltic Dry Index Forecast Using Machine Learning Methods and SHAP Explanations,” *PLOS ONE*, vol. 20, no. 3, 2025. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0325106>
- [29] S. Munkhdalai et al., “An Empirical Comparison of Machine Learning Approaches for Credit Risk Assessment,” *IEEE Access*, vol. 7, pp. 133–152, 2019. Available: <https://ieeexplore.ieee.org/document/8852736>
- [30] B. Hossin and M. N. Sulaiman, “A Review on Evaluation Metrics for Data Classification Evaluations,” *International Journal of Data Mining*, vol. 5, no. 2, pp. 1–11, 2015. Available: <https://airconline.com/ijdkp/V5N2/5215ijdkp01.pdf>
- [31] W. Li et al., “Explainable Prediction of Loan Default Based on Machine Learning Models,” *Data Science and Management*, vol. 5, no. 3, pp. 123–133, 2022. Available: <https://www.sciencedirect.com/science/article/pii/S2666764923000218>
- [32] C. Amponsah et al., “Ensemble-Based Machine Learning Algorithm for Loan Default Risk Prediction,” *Mathematics*, vol. 12, no. 21, 2024. Available: <https://www.mdpi.com/2227-7390/12/21/3423>