Bankruptcy Prediction and Analysis

June 24, 2025

SDG 9: Industry, Innovation, and Infrastructure

Group: PP-Y10-04

Submitted To: **Dr. Muhammad Ali Chauhan**

Course: Professional Practices (Y10)

Contributors:

Abdullah Imran (F2021105135)

Muhammad Affan(F2023105251)

Eman Ali (F2022105118)

Shanzy Abid (F2021408030)

Ahmer Nazeer (F2021408020)

Table of Contents

Team Contribution Summary

Member Name	Work Done	Percentage of 100
Abdullah Imran	Idea proposal, project development, execution, and report writing (project implementation, rejected ideas)	30%
Shanzy Abid	Idea proposal, Meeting Scheduling, project structuring, report writing (introduction)	20%
Eman Ali	Idea proposal, engaging the stakeholders, pitching the idea to the company, and report writing (idea pitching, conclusion)	22%
Malik Affan	Idea proposal, Detailed 5W & 1H analysis, budgeting of the project, and report writing (5W, 1H)	21%
Ahmer Nazeer	Idea proposal, highlighting the pros and cons of each idea	7%

Executive Summary

For our Professional Practices assignment, we decided to pursue a real-world issue that mapped to **SDG 9:** Industry, Innovation, and Infrastructure. We picked this particular Sustainable Development Goal for a very straightforward reason: our solution seeks to enhance financial resilience within industries, particularly small and medium enterprises. SDG 9 discusses the establishment of strong infrastructure and innovation, and we thought that the launching of an AI-powered bankruptcy forecast system squarely fell within that mandate. In a global economy where companies go out of business unexpectedly, our application provides advance notice, allowing corporations to steer clear of financial woes before they reach a point of no return.

"TriCast AI – Bankruptcy Prediction" was not only a technical challenge but a complete end-to-end solution. From pre-processing of data, model selection, and fine-tuning several machine learning algorithms such as XGBoost and LSTM, to hosting it on a real-time web interface using Hugging Face Spaces, it was convoluted and multi-layered. Each step required meticulous planning, technical expertise, and a good understanding of finance as well as AI. With zero budget, we were still able to develop a functional system that could help banks, investors, and businesses themselves.

The collaboration was one of the key reasons this project worked. Each contributed according to strengths, some did the model training and back-end, others took the lead on documentation, pitching, and front-end work. Affan joined later in the project, but soon caught up with normal briefings and open communication. We divided into technical and outreach groups to manage both the development and external feedback process, and it paid dividends. Our idea was pitched to Pak Laser Engrave (Pvt.) Ltd., where we received strong validation. The company management not only appreciated the tool's relevance but also gave actionable suggestions to improve it, especially around workforce planning and data security.

Overall, the report outlines a highly detailed, technically sound system built under tight timelines. What makes it significant isn't the models, but the effect, how it serves SDG 9, makes smart financial planning available, and was vetted by an actual business. The ultimate product wasn't merely a classroom assignment but a fully functional prototype, deployed, tested, and documented with real-world applications. It demonstrates sound technical foundation, strategic vision, and the type of collaboration that brings ideas to working solutions.

1. Introduction & Chosen Project

1.1 Introduction

In today’s rapidly changing business world, companies can face unexpected financial troubles. When a business goes bankrupt, it not only affects the company itself, but also its employees, suppliers, investors, and even the economy. This is why it’s very important to identify companies at risk of bankruptcy before it’s too late. Our project focuses on creating a smart system that can predict bankruptcy risk using financial data and artificial intelligence (AI).

The impact of a bankruptcy goes far beyond the company itself. It affects a wide network of stakeholders, including employees who may lose their jobs, suppliers who may lose major contracts, investors who may suffer financial losses, and even governments that must deal with the economic consequences. That’s why predicting financial risk in advance is not just helpful — it’s essential. Our project addresses this problem directly by building a smart, data-driven system that predicts bankruptcy risk based on a company’s financial performance.

We began this project with a simple question: What if we could build an intelligent assistant that reads financial statements and warns us before a company fails? Instead of waiting for legal announcements, credit downgrades, or public news, this system would quietly scan the company’s balance sheet, income statement, and cash flow records, and detect subtle warning signs of financial distress. The goal is to create an early-warning tool that helps people take action before it’s too late.

We started by imagining a simple but powerful idea: what if we could build a tool that looks at a company’s financial reports—like balance sheets, income statements, and cash flow—and gives a clear warning if something is going wrong? Instead of waiting for a company to officially announce a problem, our system quietly checks their numbers and highlights risk early. This gives time for investors, banks, and even the companies themselves to take action.

To train our model, we used a real-world dataset from Kaggle called "Financial Performance Prediction". This dataset contains financial records of many companies over several quarters. We cleaned the data, removed errors, calculated useful financial ratios, and then used machine learning models to learn the patterns that usually lead to bankruptcy.

More than just a project, this system has the potential to make a meaningful contribution to financial stability and business sustainability. It provides a proactive solution to prevent losses, protect stakeholders, and promote better decision-making. Additionally, this project aligns with the United Nations Sustainable Development Goal 9, which focuses on fostering innovation and building resilient infrastructure. By using modern AI tools to address financial risks, we are encouraging smarter, more responsible practices in the business world.

Our final goal is not just to build a model but to make it usable. We plan to wrap this AI model in a simple web application hosted on Hugging Face Spaces, where users can upload their company's data (in CSV format) and get an instant bankruptcy risk score. This will be helpful for:

Our bankruptcy prediction tool helps different people in easy ways. Investors can use it to spot risky stocks and keep their money safe. Banks can check if a company is stable before giving loans, avoiding bad risks. Companies can use it to check their own financial health and fix problems early. Regulators can watch over businesses to keep the economy strong and stable.

In summary, our project brings together the power of technology, the principles of finance, and the urgency of real-world business challenges. By applying machine learning techniques to structured financial data, we can uncover hidden patterns that may signal early warning signs of bankruptcy. This not only enables companies and investors to make informed decisions but also supports broader efforts to maintain financial stability in the economy. Our system transforms large volumes of raw financial data into meaningful, actionable insights, helping to prevent potential economic crises before they escalate. It demonstrates how modern AI tools can be used not just for innovation but also for creating responsible, impactful solutions that serve businesses, communities, and society at large.

1.2 Chosen Project Idea: Corporate Bankruptcy Prediction

After considering multiple ideas, our group decided to work on Corporate Bankruptcy Prediction using machine learning. The reason we chose this idea is that it solves a real and serious problem: how to identify financially weak companies before they collapse. Many people, including investors, employees, and loan providers, are affected when a company goes bankrupt. If we can predict bankruptcy in advance, we can help reduce damage and losses.

We used a structured and step-by-step approach to complete this project:

1.2.1 Dataset Selection and Preprocessing

We used the Kaggle "Financial Performance Prediction" dataset, which contains quarterly financial data of companies. Our project uses a dataset with simple company money details, like total assets, debts (short and long-term), EBITDA, cash flow, sales, and profit. To get the data ready, we cleaned it up. We filled in any missing numbers. We took out weird numbers that didn't fit. We made all numbers work on the same scale. Then, we figured out key ratios, like debt-to-assets, current ratio, profit trends, and cash levels, to help our model work better.

1.2.2 Model Building

We used multiple models in our project:

Our project uses three simple tools to predict company bankruptcy. First, we use **XG Boost** for binary classification, which sorts companies into two groups: "At Risk" or "Not At Risk" of bankruptcy. Second, we use **XG Boost** for regression to give a numerical score showing how risky a company's finances are. Third, we use **LSTM**, a tool that looks at past financial data to predict how a company might do in the future.

We used cross-validation to measure accuracy, precision, and recall. We chose the best-performing model as our final "champion model".

1.2.3. Web Deployment

To make our project practical, we deployed the model on Hugging Face Spaces using a simple web interface built with Gradio or Streamlit. A user can upload a CSV file and click predict and get an instant risk score(At Risk or Not at Risk)

1.2.4. Why We Chose This Idea (Instead of Others)

During our brainstorming, we looked at other ideas too:

We thought about a few other project ideas but didn't choose them. First, Stock price prediction, which tries to guess future stock prices. Second, CO_2 emission prediction, which looks at how much pollution might happen later. Third, Marketing campaign analysis, which checks how well ads are working. Fourth, an IoT smart energy system, which saves energy using smart devices. Lastly, Hired Up, a job matching platform to help people find jobs or internships.

But we rejected those because:

We didn't pick the other ideas because they were too hard for us. They would take longer than our 6-week project time. Some needed expensive equipment or a lot of money. Others needed data we couldn't get. Also, big companies have already solved some of those problems. We chose corporate bankruptcy prediction because it was the best fit. It used our skills in AI, Python, and data analysis. It didn’t cost much money or need special equipment. It could be done in 6 weeks. Plus, it helps businesses stay healthy and avoid financial trouble, making a real difference.

Corporate Bankruptcy Prediction was the most suitable project for our team in terms of skills, resources, and timeline. It allowed us to apply machine learning in a meaningful way to solve a real-world financial problem. The project not only improved our technical abilities but also delivered a practical tool with real impact. By identifying financial risks early, this system can support smarter decisions and promote business stability.

2. Rejected Ideas

Throughout our planning phase, we considered several adjacent ideas, each with its appeal, but ultimately set them aside for practical reasons: tight timelines, gaps in our team’s domain expertise, budget constraints, or limited business impact. Here’s a quick look at the concepts we evaluated and why they didn’t make the final cut:

3.1 Stock Market Trend Prediction

We explored using historical S&P 500 daily prices to forecast next-day market moves. While exciting, this space is notoriously noisy and dominated by hedge funds with massive compute budgets. Our six-week window wasn’t enough to engineer the sophisticated feature sets and backtest frameworks required to achieve anything beyond random-chance accuracy. In short, the risk of overfitting noisy financial data outweighed the potential benefit for our stakeholders.

3.2 Marketing Campaign Performance Analysis

We toyed with drilling into impressions, clicks, and conversion metrics to help marketing teams optimize budgets. The challenge? Every campaign lives in its ecosystem of channels, demographics, and creative strategies. Building a one-size-fits-all solution would have demanded deep integrations with ad platforms (Facebook, Google Ads, LinkedIn) and bespoke A/B-testing interfaces, far beyond our core competency and timeline.

3.3 Sales Data for Economic Insights

A project that analyzed transaction-level sales data for forecasting and customer segmentation promised rich business intelligence. However, accessing real-world customer demographics raises privacy and compliance hurdles. We also lacked a dedicated data-warehousing setup to handle millions of records efficiently, and building that infrastructure would have eaten up most of our resources.

3.4 World Top Companies Financial Analysis

Profiling the financial health of the world's largest firms, revenue trajectories, profitability benchmarks, and market-cap trends sounded compelling for strategic investors. But gathering consistently formatted, multi-source data and managing currency conversions proved to be a massive ETL effort. Given our short sprint and the project's low novelty (several commercial platforms already exist), we chose to pass.

3.5 CO_2 Emission Forecasting

We loved the idea of training an LSTM to predict a country's next decade of carbon emissions. Unfortunately, our team had limited experience with recurrent neural networks and time-series forecasting at that scale. Plus, reliable emissions projections often require auxiliary inputs (policy changes, economic growth rates), which weren't in scope. Rather than deliver a half-baked prototype, we shelved this until we can assemble the right skill set.

3.6 Smart Energy Saver System

An IoT solution using Arduino sensors to automatically disconnect appliances when consumption spikes sounded like a fan favorite. In practice, hardware procurement, firmware development, and mobile-app integration would have blown past our Rs. 3,000 budget, and we lacked embedded-systems experts on the team. It was simply too hardware-heavy for a primarily software-focused group.

3.7 Automated Home Automation Platform

Home automation, remote control of lights, locks, and thermostats, has huge mass-market appeal, but building reliable integrations across Zigbee, Z-Wave, Wi-Fi, voice assistants, and cross-platform mobile apps is a massive undertaking. We realized early that delivering a polished, user-friendly ecosystem in our timeframe was unrealistic, so we backed off this hardware-and-cloud-orchestration behemoth.

3.8 HiredUp – Job & Internship Platform

Lastly, we sketched HiredUp, a mobile app to connect students with internships. While socially valuable, the space is already crowded with platforms like LinkedIn, Handshake, and internships.com. Crafting differentiated matching algorithms and wrangling university-employer partnerships would have required months of business development and sales outreach, far beyond our six-week sprint.

By weighing each idea against our core strengths, time budget, and the value we could realistically deliver, we stayed laser-focused on the bankruptcy prediction pipeline, the project with the clearest path to impact, the most accessible dataset, and the tightest alignment with our team's expertise.

3. Targeted SDG Clause

Why SDG 9? Because this goal isn't just a slogan, it's a framework for building the kind of infrastructure the modern world needs. SDG 9 focuses on developing resilient infrastructure, promoting inclusive and sustainable industrialization, and fostering innovation. In a global economy where credit systems are the backbone of business survival, having smarter tools to predict and prevent bankruptcies isn't a luxury; it's a necessity.

This project directly speaks to that need. We're embedding AI into financial decision-making, not just for the sake of novelty, but to genuinely enhance how institutions assess risk, manage portfolios, and support businesses before they encounter difficulties. That's innovation with real impact. Our system transforms raw financial data into intelligent insights through machine learning, anomaly detection, and revenue forecasting, helping banks and credit analysts act earlier, faster, and more responsibly.

When we align this project with SDG 9, we're making a deliberate statement: we're not just coding models; we're engineering digital infrastructure for a more resilient financial ecosystem. From the backend pipelines to the explainable dashboards, every part of this solution contributes to more robust, responsive, and sustainable finance, exactly what SDG 9 envisions.

4. Project Details (5Ws and 1H)

4.1 What

Overview

Corporate bankruptcy prediction involves assessing the probability that a company will become insolvent, unable to pay its debts, and may file for bankruptcy. This process is crucial for identifying financial distress early, enabling stakeholders to take action before legal or credit downgrades occur.

As a student team, our objective in conducting this project was to provide an early warning system to industries, especially small and medium enterprises, so they can assess whether they might be heading toward bankruptcy. We wanted to highlight the importance of proactive financial monitoring and data-driven decision-making.

Key Components

Our corporate bankruptcy prediction model was built using the **Kaggle "Bankruptcy Prediction" dataset**, which includes both structured financial data and temporal trends. The following components were crucial in designing our system:

Financial Indicators

Our project uses easy financial ratios to check if a company is doing well. Liquidity ratios, like the Current Ratio and the Quick Ratio, show whether a company can pay its short-term bills. Profitability ratios, like Return on Assets and Net Profit Margin, show how much money a company makes. Leverage ratios, like Debt-to-Equity and Interest Coverage Ratio, show how risky a company's debts are.

Macroeconomic Features: External indicators such as inflation, interest rates, and GDP trends to contextualize firm-level data.

Temporal and Behavioral Indicators: Patterns in revenue, cash flow, and debt movements over recent quarters to capture momentum-based warning signs.

Methodologies:

We approached bankruptcy prediction using both traditional financial modeling and modern machine learning techniques:

Traditional Models

Our project uses two simple traditional tools to check if a company might go bankrupt. Altman's Z-Score is an old method that mixes financial numbers to see if a company is at risk of failing.

Logistic Regression is a basic tool we used to sort companies into two groups: risky or not risky, and it’s easy to understand.

Machine Learning Models

Our project uses three easy tools to predict if a company might go bankrupt. The XGBoost Classifier sorts companies into two groups: “At Risk” or “Not At Risk” of bankruptcy. The XGBoost Regressor gives a numerical score to show how likely a company is to fail. The LSTM, a special kind of neural network, looks at past financial data to spot trends and predict future money problems.

Output Types:

The output of our system is simple it gives binary output (At Risk or Not at Risk). It also makes anomaly prediction. Plus, it gives time-aware predictions to show how a company’s money situation might change in the future

5.2 When

Our project timeline was 6 weeks and was completed within that time frame. The idea was proposed by our team leader, Abdullah, and after several team meetings and brainstorming sessions, we decided to implement it as a real-world solution.

Historical Development

Bankruptcy prediction has grown over time. In the 1960s, Edward Altman created the Z-score model in 1968, a simple way to predict if a company might fail using numbers. In the 1980s and 1990s, better computers helped improve these predictions with new math-based models. From the 2000s until now, machine learning and AI have made predictions more accurate by finding complex patterns. Recently, after the 2008 financial crisis and COVID-19, people became more interested in tools that can quickly spot financial risks**.**

Critical Timing

Our bankruptcy prediction tool works best when used 1 to 3 quarters before a company faces money troubles. It is especially helpful during tough economic times when more companies struggle. For the best results, we suggest checking a company’s financial health every quarter.

5.3 Where

We applied our model in a real-world industrial context by pitching it to a local company named **Pak Laser Engrave (Pvt.) Ltd.** This allowed us to test its relevance and get industry feedback. The company appreciated the potential of our model and provided valuable suggestions, which we incorporated into our development.

Global and Sectoral Use

Our bankruptcy prediction tool is useful in different markets. In developed markets, it uses detailed financial data to build strong models. In emerging markets, it helps keep economies steady even when data is limited.

Industries:

Our bankruptcy prediction tool helps different industries in simple ways. In the financial sector, it checks the risk of loans not being paid back. It watches for changes in how customers shop. In manufacturing, it focuses on companies with big investments and lots of debt.

Emerging Trends

Our bankruptcy prediction tool can grow with new trends. ESG integration adds environmental, social, and governance factors to make predictions better. Real-time analytics lets users see instant risk results on dashboards or web tools.

5.4 Who

Key Actors

Each member played a unique role in bringing the project to life. Affan joined the team during the mid-phase of development, after which Abdullah briefed him in detail about the ongoing progress, goals, and methodology so he could integrate smoothly into the workflow.

To ensure efficiency, the team was divided into two subgroups: One focused on technical research and model development. The other engaged with industry professionals for real-world feedback and validation.

Development and Research

Our project got help from different groups. Financial analysts helped us build models using money ratios. Academics tested our methods to make sure they worked well. Tech providers gave us Python, the main tool we used to create and run our system.

5.5 Why

Our primary motivation for this project was to help industries identify financial warning signs early, especially small companies that may lack advanced financial monitoring systems. We aimed to build awareness and provide tools that could be used even with limited resourcesWe chose Pak Laser Engrave (Pvt.) Ltd. for our project because it is a well-known and successful company in printing and engraving in Pakistan. Its strong reputation for quality and performance made it a good fit for testing our AI model that predicts the risk of bankruptcy.

Our goal was not just to learn, but also to create a tool that works in real life. By working with a real company like Pak Laser Engrave, we were able to test our model on real financial data, such as assets, debts, and cash flow.

The model worked well and was able to find financial risks, showing that it could help other companies, too. We also received useful feedback from the company's management, which helped us improve our model to better meet business needs.

Economic and Financial Impact

Taking early action helps save jobs, protect investments, and keep businesses running smoothly. It also prevents a chain reaction where the failure of one company causes others to fail too. This approach supports smarter lending decisions and makes financial information clearer and more trustworthy.

Strategic Value

Early warning systems help companies change their strategies before they go bankrupt. They also help international businesses handle risks in different regions. Additionally, they provide useful information for government policymakers and regulators to make better rules and decisions.

5.6 How: Methodologies and Processes

Data Collection and Preprocessing

We used data from balance sheets, income statements, cash flow statements, and stock performance. To improve the model, we created important features like financial ratios, included economic data, and analyzed trends. We also cleaned the data by filling in missing values, fixing format issues, and handling unusual or extreme values.

Model Development

After preparing our data and selecting useful features, we began training our models in the third week of the project. First, we used basic models like **logistic regression** and **decision trees** to set a performance baseline. These models helped us understand how well simple methods could predict bankruptcy. We used five-fold cross-validation to test these models properly, checking results with scores like accuracy, recall, and ROC-AUC.

Using Advanced Models

After that, we moved to more powerful techniques:

We used three main models in our project. The **XGBoost Classifier** was used to check if a company is at risk of bankruptcy or not. The **XGBoost Regressor** gave a risk score between 0 and 1 to show how high the risk is. The **LSTM (Long Short-Term Memory) model** looked at financial data over time to find patterns from the past that could help predict future financial problems.

Feature Testing (Ablation Studies)

We also tested which types of data were most useful by removing certain features (like liquidity ratios or economic indicators) and checking how much the model's accuracy changed. This helped us focus on the most important inputs.

Final Models

At the end of week 4, we picked our best-performing models:

We used XGBoost for both classification (to check if a company is at risk) and regression (to give a risk score). We used LSTM to predict future risks based on past financial data over time. After training these models with all our data, we saved them so they can be used again in the future**.**

Deployment Platforms:

Our bankruptcy prediction model is designed to be flexible and accessible. It can be integrated into internal corporate systems or hosted on web-based platforms such as Hugging Face Spaces, making it usable by both technical and non-technical stakeholders.

Process:
The process is straightforward and user-friendly:

Users upload financial data in CSV format (including balance sheet, income statement, and cash flow figures). The model analyzes the data and returns:

We performed an anomaly analysis to spot any unusual financial patterns in the company. Then, we used a binary classification to label the company as either "At Risk" or "Not At Risk." We also did a time series analysis to study trends over time and predict the company's financial status for the next quarter.

The model's results are then plotted on the graph for a better understanding of the model's prediction.

Project Budget and Tools

This project was entirely completed through the application of our ML, logic, data analytics, and communication skills, demonstrating our ability to achieve significant results without a budget.

5. Idea Implementation

Bankruptcy forecasting is at the intersection of finance, risk management, and machine learning. For lenders, banks, investors, and regulators, knowing which firms are likely to default or fail is vital: it determines lending, drives capital allocation, and enables regulatory compliance under systems such as Basel. This project provides a multi-model AI solution tightly integrating three complementary methods—a binary classifier for explicit risk prediction, an anomaly detection pipeline for identifying unusual financial profiles, and a time-series forecaster for revenue trends—so stakeholders receive early warning signs of distress. This necessitates profound experience in data cleaning, exploratory analysis, feature engineering, model training and tuning, interpretability, and deployment engineering. In the discussion that follows, we describe each step in a manner that emphasizes technical sophistication without overspecifying for business-minded readers, while keeping the discussion firmly rooted in the banking and risk-management context.

5.1 Data Cleaning and Initial Preparation

Good-quality financial information forms the foundation of sound bankruptcy forecasting. Raw data frequently comes as CSVs or spreadsheets with balance-sheet items, income-statement amounts, cash-flow measures, financial ratios, and metadata like industry, region, and company size, plus quarterly revenue series. The primary thing is to remove any fields that might leak future results or are not relevant—internal identifiers, already calculated risk flags, or anomaly scores. This keeps models generalizing to novel companies instead of capturing artifacts. Then we address missing values: financial analysts and data experts evaluate patterns of missingness, understanding that some missing ratios indicate non-reporting or true zeros. If a material ratio is missing in only a small percentage of records, we can use industry-median imputation; if missing pervasively, we reassess inclusion or search for proxies. For quarterly revenue series, we impute missing quarters when there are sufficient valid data points; otherwise, we replace with sensible fallbacks, e.g., industry averages, flagging decreased forecast reliability. We log throughout the extent and manner of imputation so stakeholders and auditors know data quality conversions. We also impose type consistency, coercing numeric fields that come in as strings and catching conversion exceptions, marking outliers—wild fluctuations in assets or liabilities—because such outliers can reflect data errors or worse problems, such as possible fraud. Lastly, we match the purified dataset to the feature sets anticipated by each model: defaults in missing features are filled out in alignment with training, additional columns are skipped, and time-series columns are isolated for forecasting. This matching process, although transparent to non-technical users, is essential to maintaining inference consistency and avoiding runtime errors in production.

5.2 Data Preprocessing and Feature Engineering

Converting cleaned data into significant features is an area where data science skills take center stage. For classification and anomaly detection, raw financial columns tend to be augmented with derived variables: trend indicators for quarter-over-quarter changes, leverage measures aggregating balance-sheet items into ratios, liquidity measures, profitability surrogates, and encoded categorical data like industry or region. First, we stick to the set of features defined during model development, but document opportunities for future improvement so that domain experts can propose further features or alternate data sources. While tree-based models such as XGBoost are quite robust against feature scaling, our anomaly detection pipeline frequently consists of scaling operations (for instance, standardization or dimensionality reduction) to efficiently identify deviations. For forecasting revenues, MinMax scaling of sequences is also a requirement to stabilize LSTM training. We take great care to reuse exactly fitted scalers and encoders during training during inference to ensure consistency. Time-series preparation also requires caution: sequences of lengths less than the desired length could be discarded or padded with careful defaults, but we always warn users of decreased accuracy under such conditions. Feature versioning and extensive documentation of transformation logic make auditability possible: as reporting standards change and new metrics are added, the preprocessing pipeline can be augmented modularly, and version control keeps track of just which transformations saw each dataset. For stakeholders in business, this phase means "the system automatically generates consistent inputs so models pick up significant risk signals," while technical audiences admire the attention in encoding, scaling, and aligning features to model expectations stored.

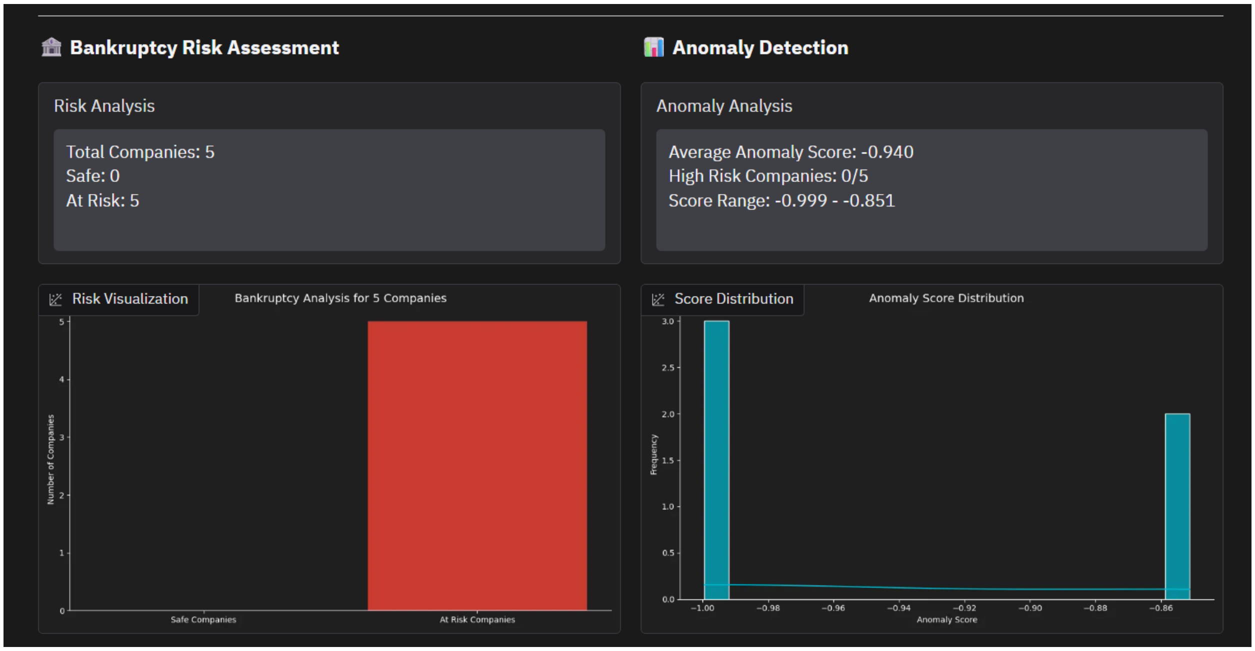
5.3 Exploratory Data Analysis (EDA)

Exploratory analysis fills the gap between raw data and modeling, revealing patterns that guide feature selection and model plans. We begin by calculating descriptive statistics—means, medians, ranges, and shapes of distributions—for each financial measure, pitting companies that would eventually go bankrupt against their healthy counterparts. Comparisons frequently yield stark signals: for instance, failing companies will have unusually elevated leverage ratios or shrinking profitability margins. Visual inspections of missingness shed light on whether particular industries or firm sizes have missing reports; if, for instance, small companies often leave out specific measures, we are aware that model confidence in those instances will be lower and readjust expectations accordingly. Correlation checks between features reveal multicollinearity—e.g., high correlations between total assets and total liabilities—which, although acceptable to tree-based models, indicate aggregation of variables into ratios for easier interpretability. We also measure class imbalance, usually discovering that bankrupt occurrences are infrequent; this knowledge contributes to directed strategies during model training to avoid trivial "always predict safe" solutions. When analyzing revenue time series, plotting sample paths reveals that firms on their way towards default normally exhibit persistent decreases or unstable fluctuations. Such findings form the rationale for incorporating an LSTM predictor. We use the anomaly detection model on historical data to plot score distributions for healthy and distressed companies to inform threshold choice (e.g., top quartile) that optimizes early warning versus alert volume manageability. Throughout EDA, data scientists and domain experts work together: statistical results are communicated as business stories ("here's why a higher debt-to-equity ratio is important"), establishing trust before model training begins. The records of EDA transformations, visualizations, and interpretations also benefit auditors in regulated bank environments, assuring that model building is based on solid data insights.

5.4 Model Training

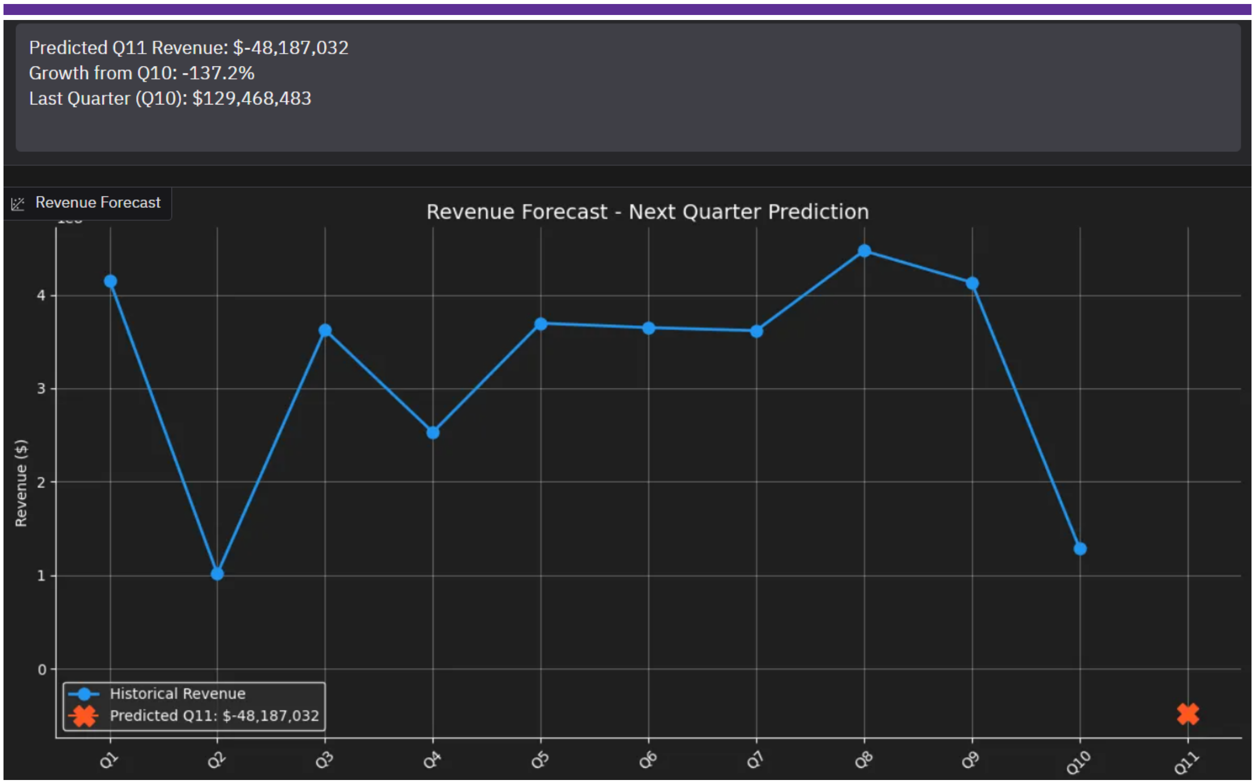
Training three complementary models requires thorough experimentation, calibration, and verification, each targeting unique aspects of bankruptcy risk. The first one is an XGBoost classifier that was trained on labeled historical data showing whether firms defaulted over a fixed horizon. Data scientists address class imbalance by modifying sampling or XGBoost's weighting factors so that the model learns default case signals even for rare defaults. Hyperparameter tuning using cross-validation optimizes tree depth, learning rate, and subsampling to achieve maximum AUC-ROC for overall discrimination, but more importantly, recall for defaults at acceptable precision so that analysts are not flooded with false alarms. From training, we obtain feature importances and calculate SHAP values for interpreting individual predictions—a hard requirement in banking, where credit officers need to know why a borrower is alerted. Cross-validation over periods and industries assures robustness; in case of a decline in performance in some sectors, we take into account sector-specific normalization or extra features. The final model, stored with metadata specifying feature ordering, hyperparameters, and performance measures, facilitates reproducibility and audit readiness.

The second pipeline is an XGBoost-based anomaly detection pipeline. Instead of predicting bankruptcy outright, it trains a "normal" financial signature—maybe by predicting anticipated next-quarter revenue or some other health indicator—and flags deviations as anomaly scores. The pipeline consists of preprocessing steps such as scaling or dimensionality reduction, followed by regression. Data scientists thoughtfully craft these transformations to capture small deviations without blowing noise up. Thresholds for anomaly flagging are determined using EDA and checked against past bankruptcy occurrences, providing an independent early warning. Explainability is still important: we create feature contributions that indicate which measures look anomalous compared to predictions. The pipeline, versioned and documented, permits tracking score distributions through time; abrupt changes can indicate economic distress or data problems.



Risk Analysis and Anomaly Detection prediction

The third model is an LSTM for time-series forecasting of revenues. As we understand that falling revenue usually precedes distress, we train the LSTM on histories of past quarterly revenues, cleaned and scaled, to forecast the next quarter. Interpolation techniques by time-series specialists help keep real trends intact and prevent distortions. The architecture, unit-tuned, dropout-adjusted, and learning rate-tuned, gets trained with early stopping to achieve a balance between fit and generalization. Evaluation is centered on not just the average error of forecasts but on sequences that result in recognized bankruptcies, ensuring the model accurately forecasts collapses in a timely enough manner to be actionable. Projected rates of growth then inform risk calculation, augmenting the classifier and anomaly detection. Everything produced by models—trained weights, scalers, transformation pipelines—is stored with full metadata to allow for reproducibility and auditability.



Time Series Forecasting of the next quarter prediction

During model training, we exercise strict version control and logging. Experiments are documented by data scientists, random seeds and library versions are logged, and evaluation reports are generated that inform governance processes. Coordination with credit analysts ensures that feature selections and model results are appropriate from a domain perspective. This symbiosis of sophisticated machine learning skills and domain collaboration forms the basis of a strong bankruptcy prediction capability.

5.5 Model Testing and Evaluation

Reliable AI requires thorough testing from various perspectives. On held-out test sets, for the classifier, we report AUC-ROC and precision-recall curves but, in particular, emphasize recall on the bankrupt class at thresholds that are tolerable to business teams. We also calibrate output probabilities to ensure reported risk scores meaningfully reflect observed default rates, enabling decision-makers to set risk-based policies. Robustness tests vary across time horizons and sectors, where performance is weak in some classes, we record constraints and weigh improvements. The anomaly detection process is verified through score distribution tests for healthy vs. distressed companies and backtesting: running the model on past data to test lead time before default, demonstrated in case studies, so non-technical stakeholders can view tangible instances. For revenue forecasting, we quantify forecasting errors both in general and in recession scenarios. Visualizing forecasted vs. actual revenue paths, for example, companies allow both technical teams to validate model quality and business users to understand how trend forecasts arise. We also perform a sensitivity analysis on missing or noisy inputs to identify data quality requirements.

Notably, we evaluate combined signals: combining classifier probability, anomaly score, and forecasted trend for integrated risk assessment. Data scientists consider straightforward rule-based blends and meta-model ensembles, measuring gains in early detection against possible rises in false positives. This analysis guides business decisions regarding reasonable trade-offs between catching more risky cases and the burden of reviewing additional alerts. Stress testing under simulated stressful economic conditions guarantees models continue to be useful in times of downturn. All steps of the evaluation, findings, assumptions, and constraints are carefully recorded, both generating summary insights for the business audience and full logs for technical teams and auditors. This open evaluation process instills trust that the AI system will work well in production banking scenarios.

Model	Testing Score
XGBoost Classifier	0.9877 Accuracy (99% correct)
XGBoost Regressor	0.195 RMSE (81.5% correct)
LSTM Time Series	0.0827 MAE (91.73% correct)

5.6 Deployment and Business Integration

Transitioning from prototypes to a user-facing application requires a lot of engineering and governance effort to envelop models with a UI. We provide a dashboard-like experience where analysts or credit officers load financial information and get an in-depth risk report: probability of classification, anomaly summary, and revenue forecast trend. Business users have an intuitive experience—create a CSV, hit "Analyze," and get transparent summary views like "High risk: 78% probability; anomaly score exceeds threshold; forecast says 12% revenue decrease next quarter"—accompanied by charts that display these results without having to know the underlying technical terminology. In the background, models and preprocessing utilities are loaded during startup for speedy inference. Data uploaded for processing gets cleaned and aligned automatically; models execute sequentially or in parallel; output is formatted into readable text and graphics.

Engineering skill supports this pipeline to be resilient, process edge cases, and log each step of processing for audit trails.

Scaling and performance issues are dealt with upfront: tree-based model inference is still quick for up to hundreds of records, whereas LSTM forecasting can be batched when portfolios become large. We track end-to-end latency to ensure a timely response. A small server will suffice for demonstration or internal pilot deployments; if usage increases, we anticipate containerization (e.g., Docker) and orchestration (e.g., Kubernetes) for autoscaling under heavy load. Security and compliance come first: financial information is private, so for actual deployments, we implement robust access controls, encryption in transit and at rest, and rigorous data isolation or automated destruction after processing. Integration with current banking systems includes creating APIs or connectors such that model outputs can be input into credit decision platforms, portfolio dashboards, and regulatory reporting tools. Even more crucial is explainability: per inference, we produce feature-contribution explanations so analysts see drivers of risk, meeting both decision-making requirements and regulatory audit needs. Alerting workflows are set up to alert risk teams when thresholds are hit, allowing timely human oversight. Continuous monitoring takes snapshots of input and output distributions and detects drift, and governance processes track versions of data, models, and deployments. This end-to-end integration illustrates how difficult engineering and organizational coordination transforming AI models into reliable parts of banking risk management.

5.7 Integration, Monitoring, and Maintenance

A bankruptcy prediction system must run continuously with careful integration, monitoring, and maintenance. We establish pipelines to retrain models regularly, quarterly, or when drift measures reach thresholds, consuming new data, such as realized bankruptcy events, to update model precision. Automated drift detection keeps track of distributions in incoming data compared with training distributions; large deviations initiate reviews. Where real defaults are experienced, outcome feedback loops send those labels back into training data. Performance dashboards monitor inference times, error rates, and alert volumes so that engineering teams can make sure infrastructure scales reliably and affordably. Explainability in manufacturing is preserved by saving explanation data—feature contributions, per inference, so analysts may look back at why a company was flagged. Security practices require dependencies to be updated regularly, patching for vulnerabilities, and periodic access control reviews. Documentation and user documentation change as models change: as thresholds or features shift, credit officers are given new guidance on interpretation and bounds. Governance models in banking demand regular model risk assessments; development artifacts, evaluation outcomes, and deployment logs are structured for audit to ensure the system is fit for purpose and compliant with changing regulations. This ongoing cycle of monitoring, evaluation, retraining, and governance highlights that a production-ready AI solution is never "done" but needs continuous effort by data science, engineering, risk, and compliance teams.

5.8 Limitations and Future Enhancements

With a strong multi-model framework, we do admit to inherent constraints. Past financial records might not reflect unprecedented shocks—financial crises, regulatory changes, or disruption events—and thus model outputs need to be interpreted by skilled analysts in context. Irregularly or thinly reported firms provide less credible projections; there is a need for domain specialists to know when to handle model indications warily. Models learned on data from some geographic regions or sectors will not necessarily generalize worldwide; strict testing for bias and potentially region- or sector-specific adjustments are necessary. As accounting standards change, the preprocessing and features need to be refined, requiring modular pipelines and continued interaction with subject-matter experts. Sophisticated methods—such as multivariate time-series models with cash flow and expense patterns, attention-based models, or graph-based approaches with inter-company relations—propose possible improvements but come at the cost of increased data sourcing and engineering complexity. More robust uncertainty quantification, such as including confidence intervals for forecasts and risk probabilities, would assist analysts in balancing model recommendations with due weight. Scaling to real-time or streaming data settings comes with added engineering complexities. Weaving together non-traditional data sources—news sentiment, market indicators, supply-chain signals—holds out the promise of more nuanced insights but involves the mastery of processing unstructured data. Building interactive explanation dashboards enables analysts to drill down into drivers and run scenario analyses, but it involves front-end engineering and careful UX design. Lastly, as regulatory expectations change, the system and documentation need to update rapidly. Identifying these limitations and planning for future improvements guarantees that stakeholders know both the existing capabilities and the path forward.

6. Idea Pitching

On 2 June 2025, our team visited Pak Laser Engrave (Pvt.) Ltd., a prominent printing cylinder manufacturer in Lahore, Pakistan. The purpose of our visit was to pitch an AI-based bankruptcy prediction system to the company’s management team. This tool aims to support early detection of financial risks using company financial data and help mitigate potential losses through proactive strategies.

Our solution is designed not as an academic prototype, but a real-world application aligned with SDG-9 (Industry, Innovation, and Infrastructure). It represents a quiet, behind-the-scenes assistant, analyzing company balance sheets, cash flows, and income statements — and flagging subtle signs of distress before they become crises.

6.1 Company Background & Meeting Participants

Company: Pak Laser Engrave (Pvt.) Ltd.

Pak Laser Engrave is a leader in the rotogravure printing industry, providing custom cylinder engraving solutions to packaging and printing clients. The company's focus on high precision, quality, and production efficiency makes it an ideal candidate for financial health innovation tools.

People Present During the Pitch:

1. Mr. Muhammad Asif – Head Manager: Oversees plant operations, financial decisions, and long-term strategic planning.
2. Mr. Shahwaiz Ahmad – HR Manager: Responsible for workforce planning, organizational development, and sustainability initiatives.
3. Accounts & Production Team – Represented technical and financial departments. Role: Interacted with the tool's financial data perspective and asked implementation-related questions.

The Concept:
We introduced our system as a digital financial analyst that reviews a company's quarterly financial statements and provides a bankruptcy risk score before external signs appear. Instead of reacting to red flags like credit downgrades or court orders, our system gives management the edge of predictive insight.

Technical Foundation:

We used the "Financial Performance Prediction" dataset from Kaggle, which includes details like assets, liabilities, profitability ratios, cash flows, and debt-equity ratios. We applied feature engineering to pull out the most important indicators, such as the current ratio, operating margin, and trends in net income. Several machine learning models, including Logistic Regression, Random Forest, and XGBoost, were trained and tested on new data. These models gave high accuracy and very few false alarms. To make it easy for users, we built a simple web app hosted on Hugging Face Spaces. Users can upload a CSV file with their company's quarterly data and instantly get a risk score, along with easy-to-understand visual results.

9.2 Alignment with SDG-9 and Business Benefits

SDG9-Focus:
Goal 9: Industry, Innovation and Infrastructure. This model promotes sustainable industrial development by ensuring economic resilience. Corporate failures not only impact companies but can also lead to layoffs, unpaid suppliers, and sector-wide disruptions.

We chose SDG-9 because it focuses on industry, innovation, and building strong businesses. Our project supports this goal by helping companies, especially small-scale businesses, understand their financial health. This can help them grow and avoid bankruptcy. We specifically focused on SDG 9.3, which aims to increase access to financial services and markets for small businesses. By using our AI-based risk prediction system, these businesses can make smarter decisions, get better access to loans and investments, and improve their chances of success. Our goal was to support business growth and make financial planning easier and more reliable for everyone.

Strategic Business Benefits for Pak Laser Engrave:

Our system offers several valuable benefits for companies. It acts as an early warning system by detecting financial problems before they happen, helping businesses take action in advance. By relying on real data, it supports smart and informed financial decisions instead of guesswork. This builds trust with banks, investors, and regulators by showing that the company is actively monitoring risks. The system can also be customized using the company's historical data, which improves accuracy. Best of all, it runs in the cloud, so there's no need for complex IT setups—just upload the data and get clear results instantly.

9.3 Feedback and Team Interaction

Feedback from Mr. Asif (Head Manager):

He found the tool "practically useful and timely." He related it to internal struggles they face due to sudden supplier payment issues and project delays. He expressed interest in piloting the tool using their last 3 years of financial statements.

Feedback from Mr. Shahwaiz (HR Manager):

He appreciated the innovation and inquired about employee impact forecasting, proposing that financial risk scores be tied to workforce planning to avoid unnecessary layoffs in bad quarters.

Team Questions and Suggestions:

- Can it be linked to external economic events like currency drops? Yes, with additional features (inflation, GDP indicators), we can scale the model.
- What's the risk of false alarms? We've optimized precision-recall balance to minimize both false positives and negatives.
- Is the system secure for sensitive financial data? Yes. It uses secure upload protocols and performs all processing in memory.

7. Conclusion

This project was more than just a university assignment—it was our way of solving a real problem that many businesses face. When companies go bankrupt, it can hurt not only the business owners but also workers, suppliers, investors, and the economy. That's why we built a smart tool using AI that can look at a company's financial data and warn if it's at risk of going bankrupt. It gives people a chance to fix the problem before it gets worse.

We worked hard as a team, each person doing what they were good at. Some of us trained the AI model, others worked on the website, and some focused on writing and presenting. Together, we created a working system that actually does what it's meant to do. Even though we had no money to spend, we still managed to make a tool that works.

What made this project even more special was that we didn't just keep it in the classroom—we took it to a real company, **Pak Laser Engrave**, and shared our idea with them. They really liked it and said it could be helpful in real life. They also gave us useful feedback to improve it, like how to better protect data and help in staff planning.

Our project also supports **SDG 9**, which is all about making industries stronger and smarter through innovation and good planning. By warning businesses early about financial risks, our tool helps them stay safe and grow better. That's why this project matters—it's not just about avoiding bankruptcy, it's about building a stronger future.

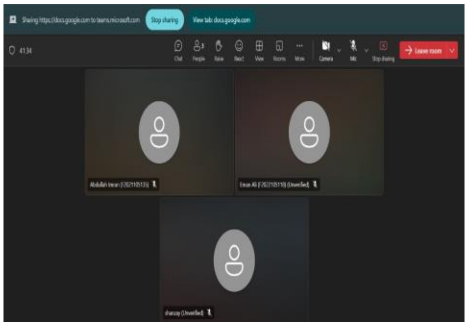
In short, we built something useful, tested it in the real world, and saw how it could make a difference. We learned a lot, worked well together, and proved that technology can really help solve big problems when used in the right way.

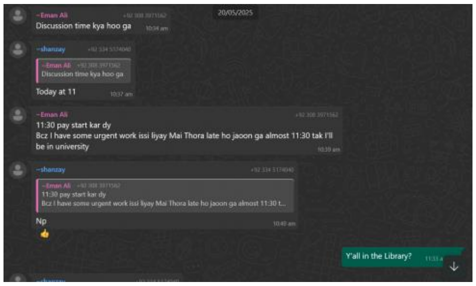
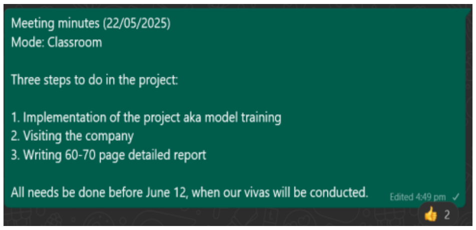
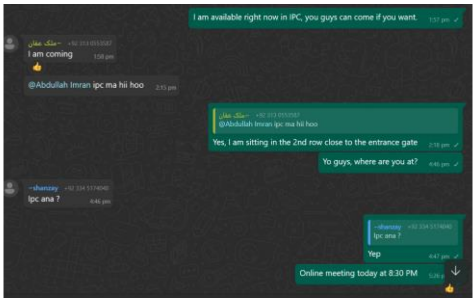
References

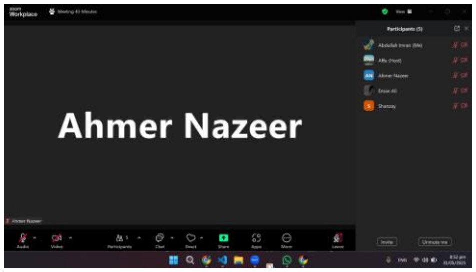
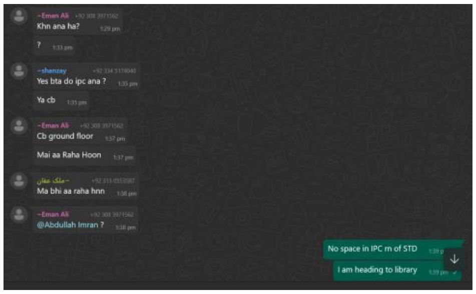
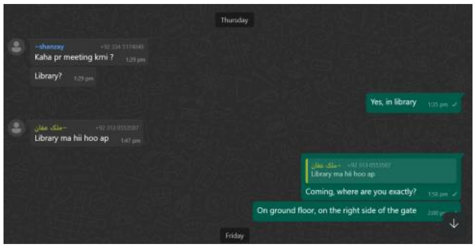
1. United Nations Sustainable Development Goals – SDG 9: https://sdgs.un.org/goals/goal9
2. Kaggle – Financial Performance Prediction Dataset: https://www.kaggle.com/datasets
3. Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. Journal of Finance.
4. Machine Learning Mastery – Brownlee, J. (2021). Predictive Modeling for Financial Distress.
5. Hugging Face Spaces – ML App Hosting Platform: https://huggingface.co/spaces
6. Company Visit, Interview Notes: Pak Laser Engrave, June 2025
7. Project Demo Link: https://huggingface.co/spaces/AbdullahImran/TriCast-AI/tree/main
8. Idea Pitch Link: https://drive.google.com/file/d/1M7tEoVlwRbo7Dt0yz6tM-nq2vACLmh0r/view

Project Evidences

Meetings conducted

No	Members	Date	Agenda	Evidence/Mode
1	Abdullah Imran, Eman Ali	19/03/2025	Organization structures, HR, centralization, decentralization	https://docs.google.com/document/d/18hoFlL6bN0R-wEZ-b4kXzCFU5fxocb8OEDDcUoYUj-Y/edit?usp=sharing
2	Abdullah Imran, Shanzy Abid, Eman Ali	14/05/2025	Team Development & Decision-Making Process	

3	Abdullah Imran, Shanzy Abid, Eman Ali, Malik Affan	20/05/2025	Discussing the flow of the project	 Mode: At CB1
4	Abdullah Imran, Shanzy Abid, Eman Ali, Malik Affan, Ahmer Nazeer	22/05/2025	Deciding on the next steps of the project	 Mode: Classroom, North Building
5	Abdullah Imran, Shanzy Abid, Eman Ali, Malik Affan	29/05/2025	Showing the project demo	 Mode: IPC
6	Abdullah Imran, Shanzy Abid	30/05/2025	Discussing the project ideas	

8	Abdullah Imran, Shanzy Abid, Eman Ali, Malik Affan, Ahmer Nazeer	31/05/2025	Discussing the ideas on how to pitch the project to the companies	
9	Abdullah Imran, Shanzy Abid, Eman Ali, Malik Affan	02/06/2025	Meeting to pitch the idea to the company	
10	Abdullah Imran, Shanzy Abid, Eman Ali, Malik Affan	17/06/2025	Dividing the presentation parts	 Mode: Library
11		19/06/2025	Presentation Rehearsal	 Mode: Library

Project Pitch

Pitch Date: 02/06/2025

Company: Pak Laser Engrave (Pvt.) Ltd.

Pitch Duration: 23 Minutes

Location: 114 Shan Rd, Quaid-e-Azam Industrial Estate, Lahore

Video: https://drive.google.com/file/d/1M7tEoVlwRbo7Dt0yz6tM-nq2vACLmh0r/view?usp=sharing