INTELLISCORE - Ai Based Credit Scoring System

This BS Project report is submitted to the Department of Computer Science as partial fulfillment of Bachelor of Science in Computer Science degree

https://github.com/Fazulsden/Fyp-FinalReport.git

by

Shahmeer Khan, Fazul Al Rehman. Nosherwan Shaikh, Abdur Rehman

Advised by

Miss Tasbiha Fatima
Lecturer
Department of Computer Science
School of Mathematics and Computer Science (SMCS)
Institute of Business Administration (IBA) Karachi

Fall 7th semester 2024
Institute of Business Administration (IBA) Karachi Pakistan

Contents

1 Contributions **1**
- 1.1 Shahmeer Khan . . . 1
- 1.2 Nosherwan Shaikh . . . 2
- 1.3 Abdur Rehman . . . 4
- 1.4 Fazul Al Rehman . . . 5

2 Software engineering life-cycle **8**

3 App walkthrough **12**

4 Project Abstract **18**
- 4.1 Background and Approach . . . 18

5 Experimental settings **20**
- 5.1 Control Group: . . . 20
- 5.2 Experimental Group: . . . 21
- 5.3 Project Evolution . . . 22
- 5.4 Software and Hardware: . . . 23
- 5.5 Parameter Settings: . . . 25
- 5.6 Data Preprocessing: . . . 25
- 5.7 Evaluation Metrics: . . . 26
- 5.8 Experimental Setup: . . . 26

6 Results and Discussion **28**
- 6.1 Implementation Highlights . . . 29
- 6.2 Validation Outcomes . . . 29
- 6.3 Model discussion based on experimentation . . . 29
- 6.4 Limitations . . . 31

7 Conclusions , Limitations and Future Work **32**
- 7.1 Limitations and issues faced: . . . 32
- 7.2 Future Work: . . . 37

7.3 Conclusion: 38

8 References and Bibliography **40**

Chapter 1

Contributions

The following section highlights a matrix of the contributions of each individual across the project.

The following tables are catered to each member's contributions, the project part defines the submitted document, the member contribution specifies the part the member worked on, the percentage defines their contribution in percent to that part.

1.1 Shahmeer Khan

Role: Project Lead

Handled the work assignment for the team and a majority of the communication overload with the industry. Being the lead i had to take up the role of distributing work at every stage, this also came with it's challenges as i had to force everyone to work, to have meetings and to contribute as that's the lead's job, at many occasions i also had to handle my member's shortcomings and make sure we deliver everything on time and in good condition which took many extra hours of effort. I was also the focal point of communications with Meezan constantly urging them to arrange meetings etc as i was the one who took up their project in the first place.

Project Part	Member Contribution	Contribution %
Proposal	-Scope -Gantt Chart -Wireframes -User Flow Diagram	100% for all of these chapters and subchapters.
SRS	-Case overview -Scope -User Interface requirements	100% for all of these chapters and subchapters.
SD	-User Interface Design and Implementation	100% for all of these chapters and subchapters.
Frontend	-Complete UI designing for the pages Reports, Help	100%
	-Worked on the UI for detail entry, and dashboard page	30%
	-Implementing Signup, Login, settings profile UI and refining previous design	50%
Backend	-Worked on backend logic for settings profile to refine it	50%
	-Implementing Signup, Login logic as well as authentication	50%
	-Worked on backend logic for form, dashboard	40%
Database	-Implemented a Postgres solution for our app	100%
Model	-Implementing the student teacher distillation logic, ensemble approach via Histgradboost	50%
Data Preparation	-Curated the Pakistan HIES dataset to match our needs	100%
Final Report	-Limitations and issues faced	100%
	-Results and discussions	20%
Report compilation	-Compiled and designed all the reports so far via Latek Overleaf	100%
OBE	-Target Users-frontend	100%

1.2 Nosherwan Shaikh

Role: Member

Handled the base logic for model designing aswell as aided in frontend development. While the rest of the team focused their attention onto the webapp, the model was implemented by

Nosherwan, starting with the initial synthetic data generation going all the way to the testing and training phase, this side of the project was held up by him, it wasnt until our ensemble approach that the lead started working on the model aswell.

Project Part	Member Contribution	Contribution %
Proposal	- Introduction -Problem Statement -Executive Summary - Features	100% for all of these chapters and subchapters.
SRS	- Problem Domain -Machine Learning -Specifications -Hyperparameters -Evaluation Metrics -Formulas -Stakeholders -Data Domain -System Functionalities -Market and Use Case -Analysis -Constraints and Assumptions -Tech Stack	100% for all of these chapters and subchapters.
SD	- Algorithms and Data Structures	100% for all of these chapters and subchapters.
Frontend	-Worked on the UI for dashboard page	50%
Backend	-Worked on backend logic for form	60%
Final Report	-Project abstract -Background and approach	100%
	-Results and discussion	10%
Model	-Designed the base model as well as trained and tested through several models to pick MLP	50%
Standee	-Full design	100%
OBE	-App progress and envisioned benefits	100%
React Frontend	-Although not used in the final product, Developed the frontend for all the login and signup pages on react for a possible shift to a separate frontend and backend system	100%

1.3 Abdur Rehman

Role: Member

Handled most of the research work and literature review of the project. While not much work went into the app side of things Abdur Rehman handled our research and literature work to the fullest where he browsed through several resources to give us a good baseline of how credit scoring works and how to go about our project.

Project Part	Member Contribution	Contribution %
Proposal	-Literature Review -Use Cases -Market Research and Analysis -Requirements	100% for all of these chapters and subchapters.
SRS	-Functional Requirements Specification (Stakeholders, Actors & Goals, Use Cases) In proposal, include: -Literature Review (Text Analytics, Computer Vision and Interpretable Machine Learning in Credit Scoring) In market research & analysis: -Market Research -Ethical Implementation -Other Industry Projects	100% for all of these chapters and subchapters.
SD	-Interaction Diagrams -Class diagram and Interface specification	100% for all of these chapters and subchapters.
Frontend	-Worked on the UI for dashboard page, sidebar functionality	20%
	-Sidebar functionality in reports page	10%
Final Report	-Results and discussion	70%
OBE	-Terrible actors- backend	100%

1.4 Fazul Al Rehman

Role: Co lead

Handled most of the communication aswell as networking with the supervisor aswell as IBA staff to get a majority of work done. An integral part of our project was keeping the supervisor aswell as the IBA faculty in the loop at all times, due to distance issues for the other members, Fazal mainly took the initiative to hold several individual meetings with Ms Tasbiha aswell as to get support from other IBA faculty on certain issues while also communicating the problems we face and the solutions we offer to the FYP committee members to ensure smooth sailing.

Project Part	Member Contribution	Contribution %
Proposal	-Requirements -Features -Work division -Implementation timeline - Data Preparation - ML Ops Integration - CI/CD Pipeline - Deployment Strategy - Ethical Considerations	100% for all of these chapters and subchapters.
SRS	- System Requirements - Functional Requirements - Non-Functional Requirements - User Interface Requirements - Stakeholders -Functional Requirements Specification - Actors and Goals	100% for all of these chapters and subchapters.
SD	- System Architecture and System Design - Architectural Style - Identifying Subsystems - Mapping Subsystems to Hardware - Persistent Data Storage - Network Protocol - Global Control Flow - Hardware Software Requirements	100% for all of these chapters and subchapters.
Frontend	-Worked on the UI for detail entry	70%
	-Designed the main UI for the app's core page template	100%
	-Defined the base UI for login and signup pages	50%

	-Designed the base UI templates for both settings and profile pages	50%
Backend	-Worked on backend logic for form	60%
	--Designed the base backend logic for both settings and profile pages	50%
Backend	- Defined the base backend logic for login and signup pages	50%
Final Report	-Experimental Settings	90%
Poster	-Full design	100%
React FrontEnd	-Although not used in the final product, Developed Frontend Pages for the microservices approach pages including the dashboard and entry form, more were to be collaborated with the team but focus was required on monolith	100%
React BackEnd	-Although not used in the final product, Developed APIs for all the pages for a feasible shift to a separate frontend and backend system to follow microservice architecture, which was changed when the requirements were changed from Meezan.	100%
OBE	-Terrible actors- frontend	100%

Chapter 2

Software engineering life-cycle

This section contains images of a document we compiled to highlight the meetings we've had, what we've covered across each sprint and Meezan's as well as our instructor's signature as approval of these meetings taking place.

Note:

Throughout the calender year we have approached Meezan to organize several meetings, however timing issues and unavailability on their end meant we could eventually only have 6 meetings in total, these however were mostly fruitful and a brief description has been offered for each one.

All other software engineering steps and documentation already exist in the previous documents i.e. proposal, SD, and SRS, which contain requirements engineering, wireframing (high and low level), gantt chart for work distribution, use cases, user flows and so on.

FYP Project sprints document

Highlights the meetings we've had with meetings across sprints, purpose of the meetings, the targets achieved and discussed during those meetings and key takeaways.

Sprint 1:

1st meeting (online)

Held at: 13th September 2024

Purpose:

- Discuss Meezan's project requirements
- Discuss our data requirements

Key takeaways:

- Given insights on how credit scoring works
- Tasked with gaining knowledge on finance operations

2nd meeting (physical)

Held at: 26th September 2024

Location: Meezan House

Purpose:

- Agree on project scope
- Gathering requirements from Meezan Bank for our project's approach

Key takeaways:

- Meezan allotted us full control of the app itself, requirements limited and specified to a credit scoring model
- Instructed us to perform deeper research on credit scoring applications and credit scoring models implemented globally and locally

Sprint 2:

3rd meeting (physical)

Held at: 20th February 2025

Location: Meezan House

Deliverables prepared: Project proposal, SRS

Purpose:

- Discuss software's specifications
- Demo wireframes
- Agree on access to data

Key takeaways:

- Discussion on credit scoring models
- Scope narrowed to credit scoring for bike loans
- Given bike loans form data to use for model's data fields
- Urged to work with synthetic data for base model generation

Sprint 3:

4th meeting (online)

Held at: 7th may 2025

Deliverables prepared: Fully functional webapp with working Ai model based off synthetic data

Purpose:

- Demo working model and webapp
- Gain Meezan's confirmation for access to data

Key takeaways:

- Webapp accepted
- Issues with the model
- Agreement on NDA and data after a scope review

5th meeting (online)

Held at: 11th may 2025

Deliverables prepared: Finalized webapp with previous model

Purpose:

- Request access to data
- Highlight need for data to meet deadlines and finalize product

Key takeaways:

- Agreement reached on development of NDA
- Scope renewed to use car loan data

6th meeting (online)

Held at: 21st may 2025

Deliverables prepared: Finalized webapp with previous model, alternatives for dataset

Purpose:

- To propose alternatives for data but enforce need for data

Key takeaways:

- NDA provided for Meezan's data

Acknowledgment:

I, **Mr. Usman Siddiqui**, representing the Shariah Compliance Department of Meezan Bank, hereby acknowledge that I have reviewed the contents of this FYP Project Sprints Document. I confirm that the meetings listed herein did in fact take place on the mentioned dates, and that the purposes, deliverables, and key takeaways summarized accurately reflect the discussions held between our team and the student project group. I agree to act as a witness to the authenticity and factual nature of this document.

Signature: Usman
Name: Mr Usman Siddiqui
Designation: Manager Retail
Date: 23rd May 2025

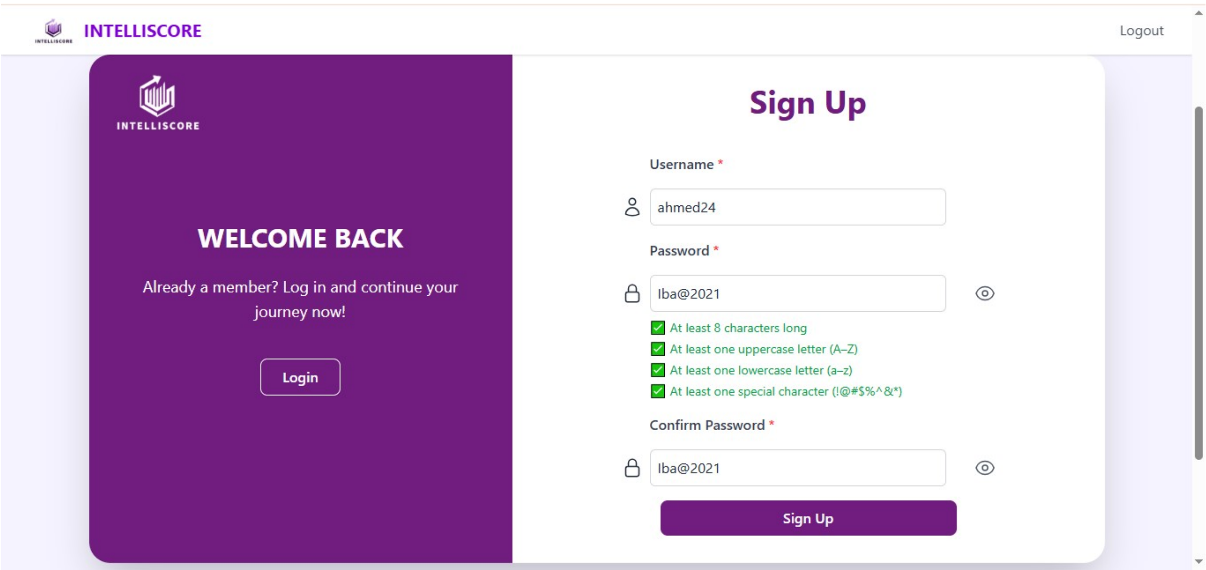
Instructor Signature: Tabiha
Instructor Name: Ms Tasbiha Fatima
Instructor Designation: Lecturer SMCS

Meetings took place.
Final report and model awaited.
Us

Chapter 3

App walkthrough

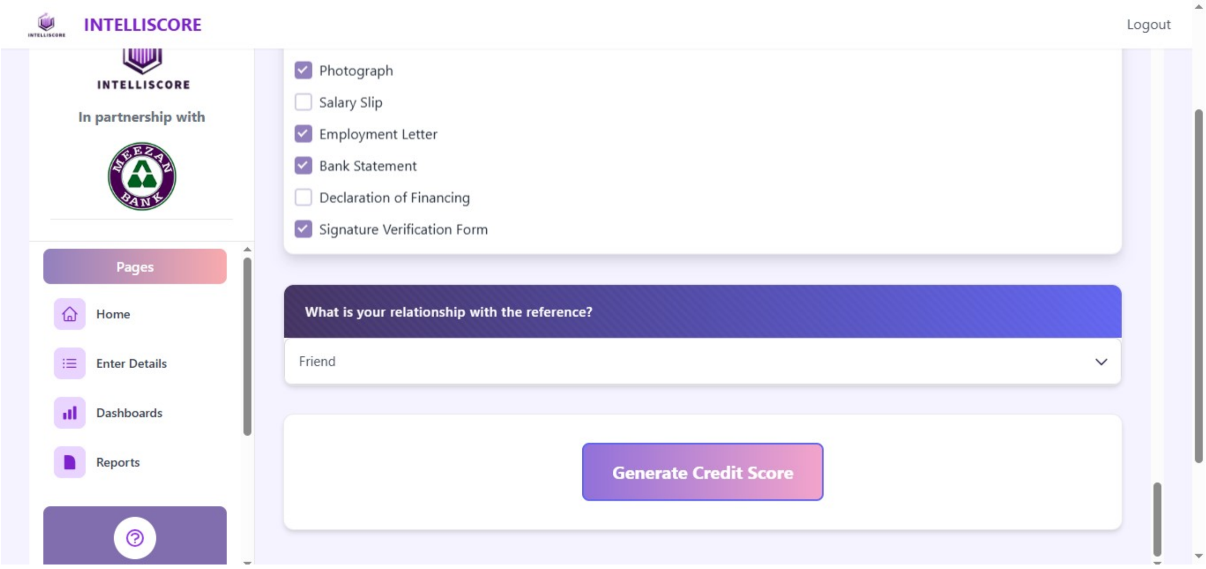
Although not an official requirement for the final report, we wished to give a user walk through for our app to highlight a user journey for our final app.



Signup Screen

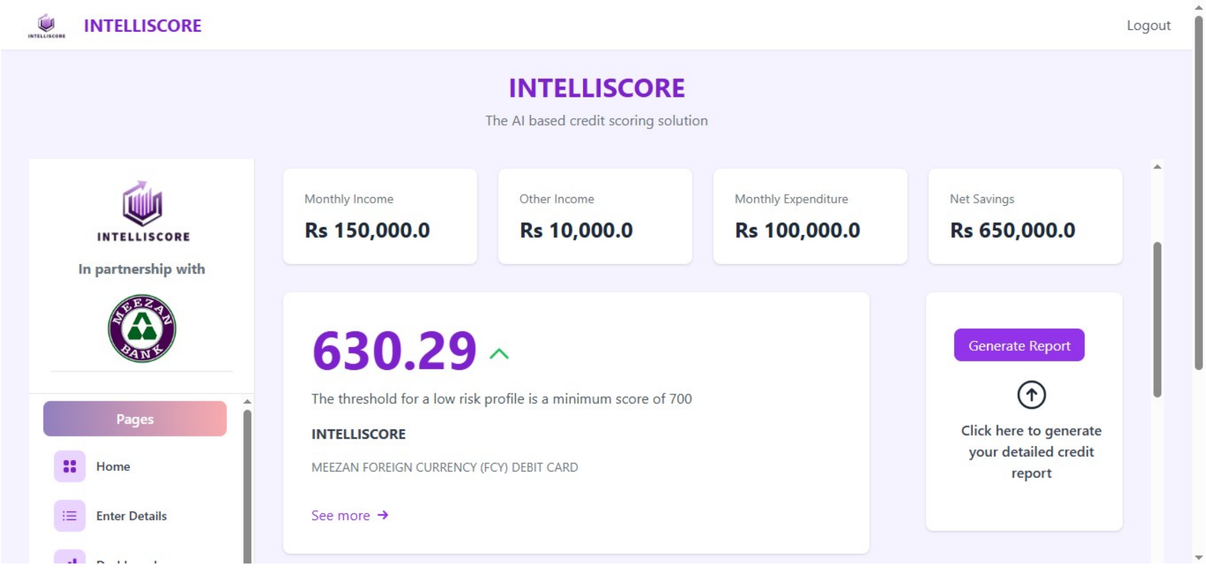
The user first signs up, ensuring credentials are valid and pass the requirement checks.

The user is navigated to home page after signing up, from here on they then navigates to the enter details page via the left side navigation bar and fill the form accordingly and clicks generate score.



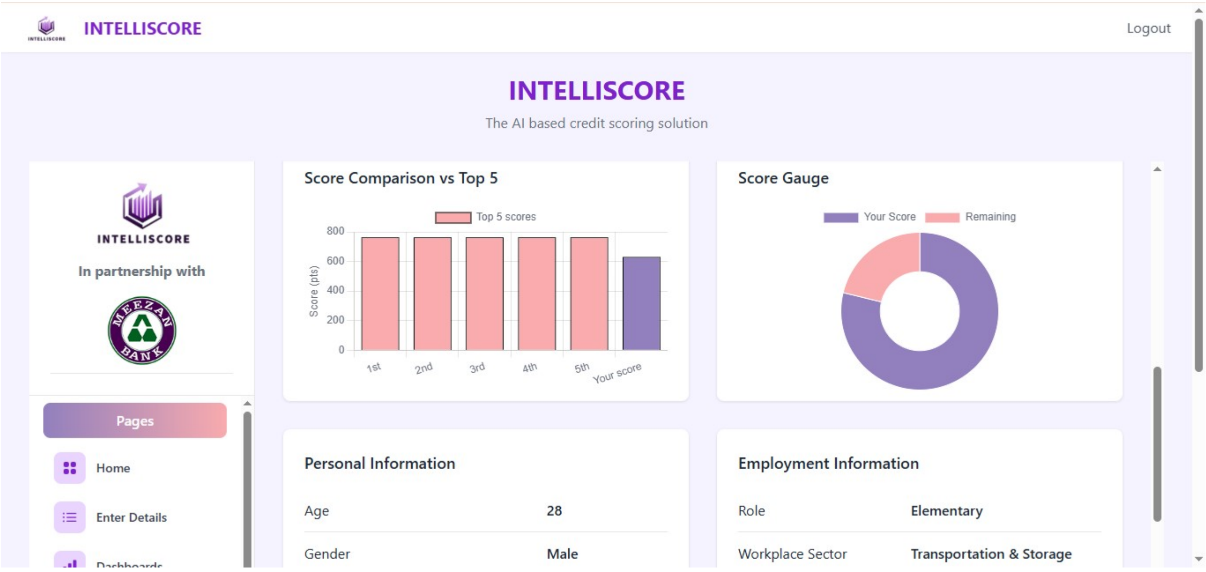
Detail Entry Screen

The user is brought to the dashboard page with their generated score as well as a few important KPIs, a generate report button can be used for reporting, explained later.



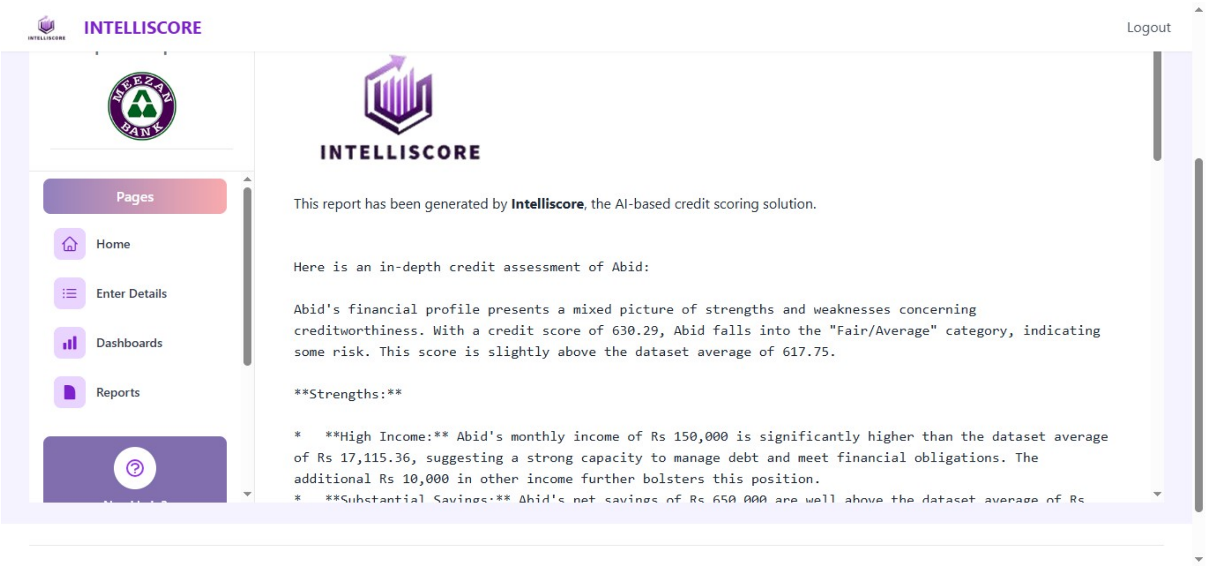
Dashboard Screen

The user can then navigate through several charts comparing their input to existing data to visualize a simpler but deeper understanding of their score.



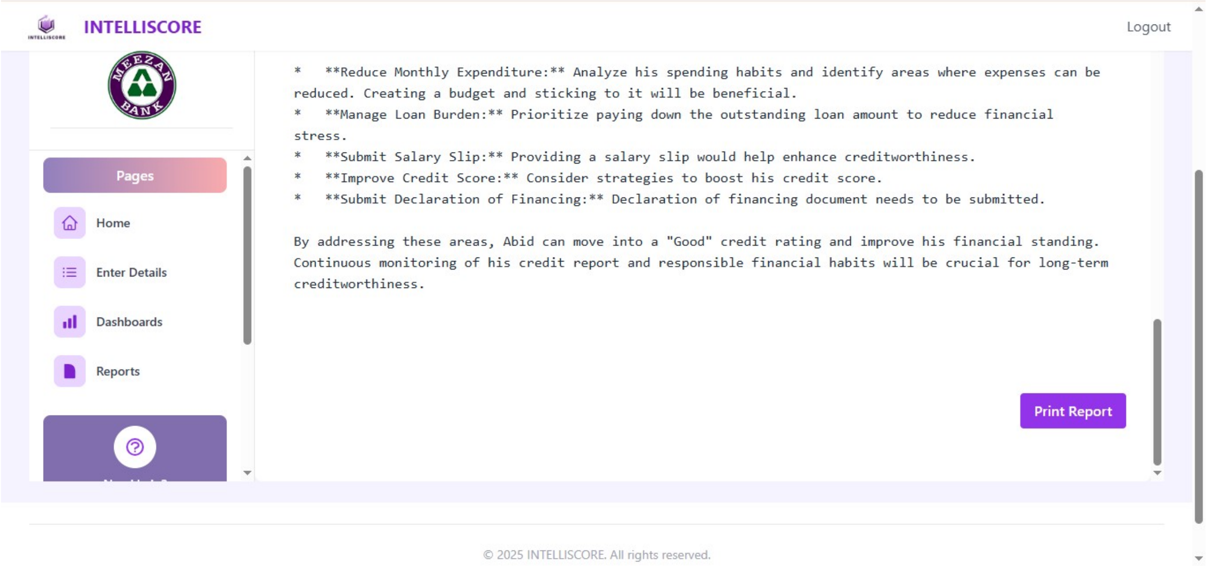
Dashboard Screen

The user can then also choose to generate a report using our API, this report gives an even deeper and more complex explanation of the score, the reason for the score and how to improve the score.



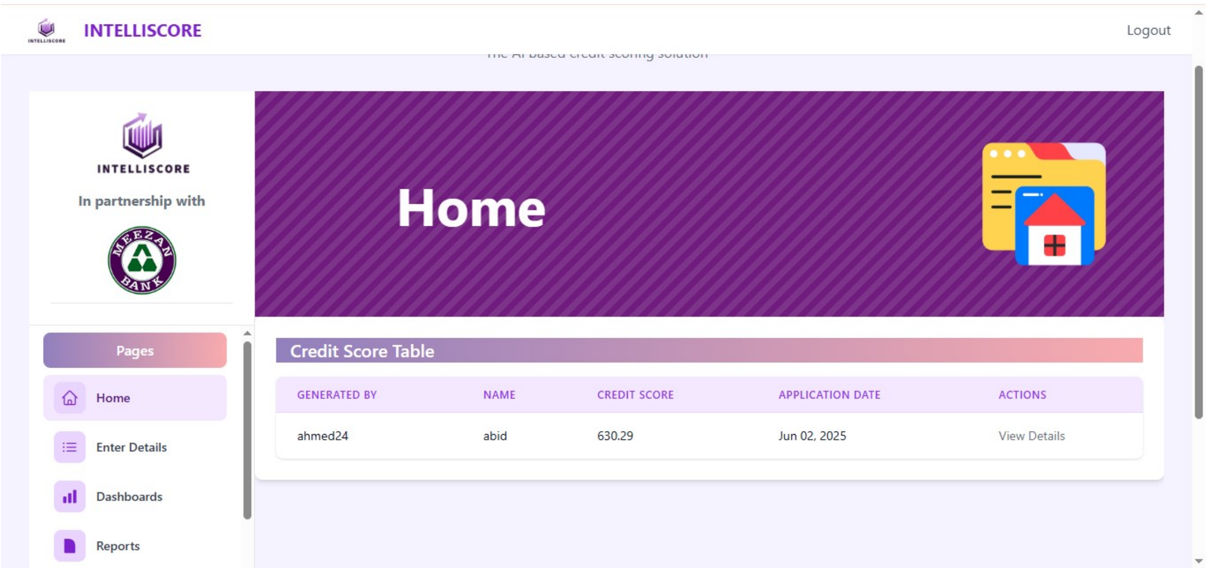
Report Screen

The user can then also choose to save or print their report if needed.



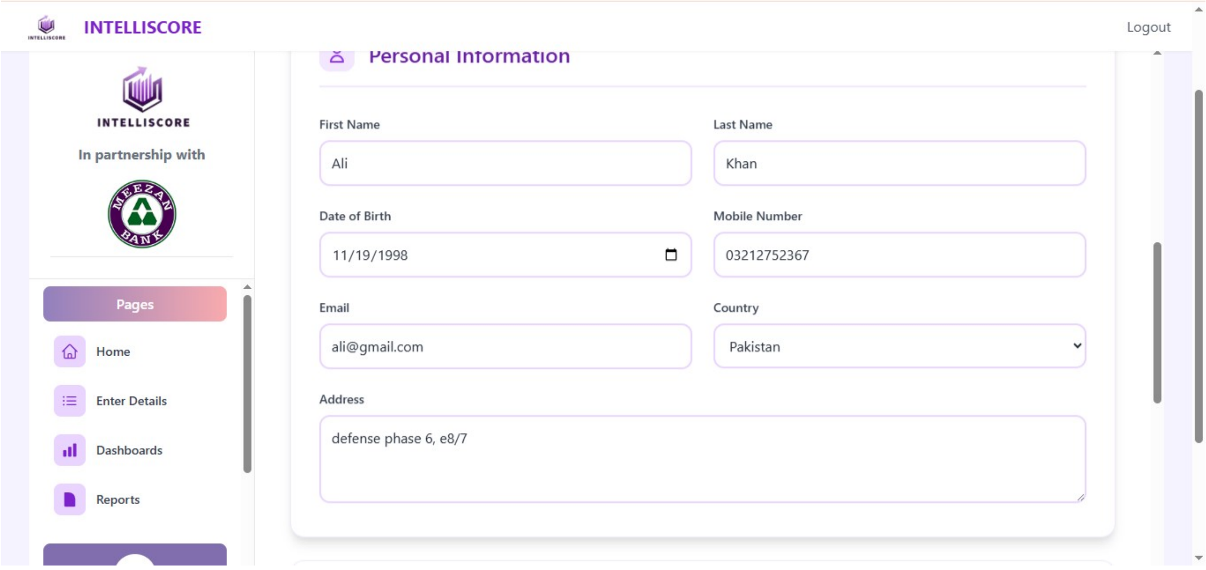
Report Screen

The score that was generated will now be updated on the home page as a record, you can click view details to go back to that score's dashboard as well.



Home Screen

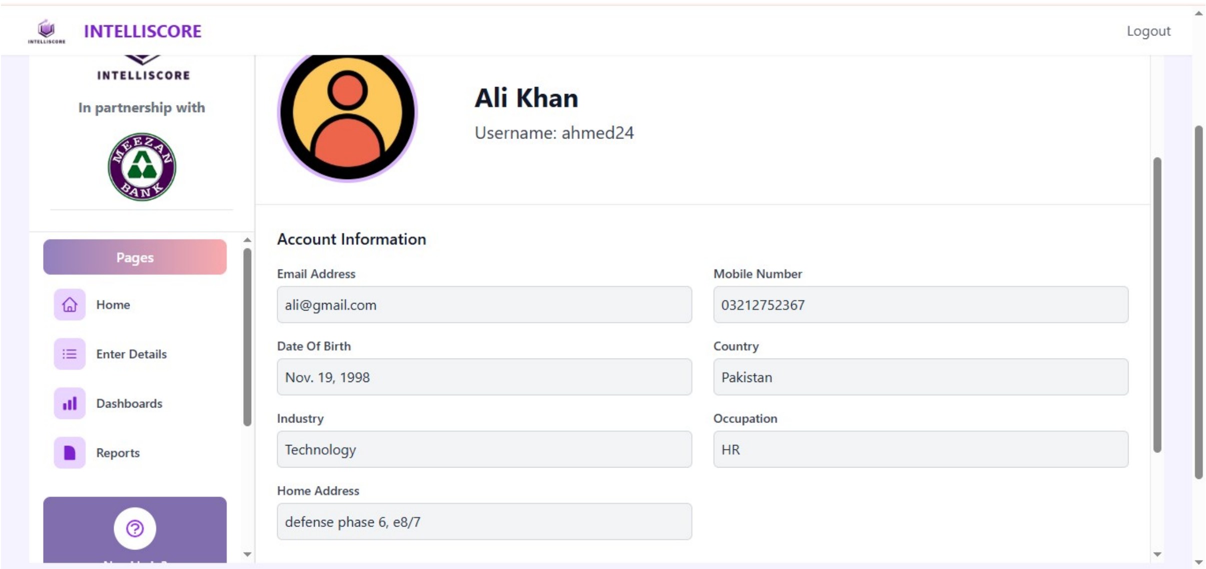
The app's main usage walkthrough ends here however there are additional features as well.



Settings Screen

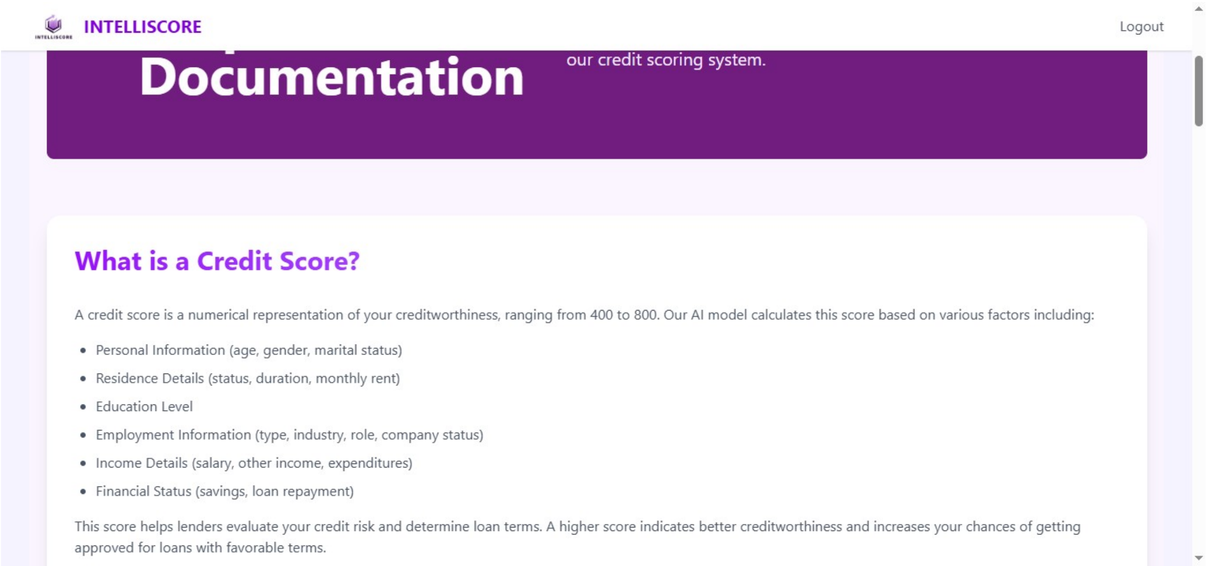
There is a settings page where the user can setup their profile accordingly, it also includes a change password tab for ease of use.

The profile changes made in the settings page can then be viewed in the profile page where they are displayed as non editable fields.



Profile Screen

Finally there's a help and documentation page to answer all your queries regarding the app, this can be accessed via the check our docs option in the left navigation panel.



Help Screen

Chapter 4

Project Abstract

This project introduces an AI-based credit scoring system developed in collaboration with Meezan Bank, aimed at enhancing the accuracy and efficiency of credit risk assessment through modern machine learning techniques. The core objective is to leverage artificial intelligence to analyze a range of financial and behavioral attributes in order to generate reliable credit scores. Recognizing the sensitivity of banking data, a formal Non-Disclosure Agreement was signed with Meezan Bank to ensure strict confidentiality and compliance during data sharing; however, despite all formalities being fulfilled, the data could not ultimately be provided by the bank. As a result, development proceeded using a high-quality dataset obtained from the Pakistan Bureau of Statistics, curated under the supervision and approval of our project advisor, Miss Tasbiha Fatima, to serve as the basis for designing and testing the credit scoring logic. The system adopts a student-teacher distillation paradigm, where the student model learns from credit scores generated by a hardcoded rule-based formula, a neural network, and a gradient boosting algorithm. This design enables the student model to generalize complex patterns effectively while leveraging the complementary strengths of each scoring approach. The current implementation successfully demonstrates the feasibility of AI in streamlining credit scoring, offering promising implications for more inclusive and data-driven financial decision-making within the banking sector.

4.1 Background and Approach

Globally, leading financial institutions such as JPMorgan Chase, Wells Fargo, and others are increasingly integrating artificial intelligence into their credit scoring systems to enhance risk evaluation, streamline decision-making processes, and broaden access to financial services. These advancements have demonstrated substantial improvements in both predictive accuracy and operational efficiency, while promoting more inclusive lending practices. Inspired by such progress, our project aims to develop an AI-powered credit scoring system tailored to the unique needs of the Pakistani banking sector.

To ensure real-world relevance, the project was initiated in collaboration with Meezan Bank, with the intent to train the model on actual customer data. A formal Non-Disclosure Agreement was signed to enable secure data sharing, following close coordination with legal and academic representatives from Meezan Bank and IBA. Despite the completion of all formalities, the bank was unable to provide the data due to internal constraints. Consequently, the development of the scoring system proceeded using a high-quality dataset sourced from the Pakistan Bureau of Statistics.

This dataset, curated and utilized under the supervision of our project advisor, Miss Tasbiha Fatima, provided a rich and contextually appropriate foundation for implementing and evaluating the model's scoring logic.

Our approach is centered around a student-teacher distillation framework. In this setup, the student model is trained using credit scores generated through three complementary methods: a hardcoded rule-based formula, a neural network, and a gradient boosting algorithm. This allows the final model to learn from both expert-defined logic and data-driven insights, enhancing its ability to generalize and predict accurately across diverse financial profiles.

To support practical deployment, we developed a comprehensive web application. The system is built with Python Django for backend processing, Tailwind CSS for responsive design, and PostgreSQL for secure data management. It features multi-user authentication and facilitates real-time credit evaluation by accepting detailed user input and generating credit scores alongside visual explanations and natural language summaries. Technical specifications and implementation details are outlined in the subsequent sections of this report.

Chapter 5

Experimental settings

To ensure rigorous testing and falsifiability in our experiments, we structured our experimental settings meticulously, including clearly defined control and experimental groups, parameter settings, software and hardware environments, and thorough data pre-processing methods.

5.1 Control Group:

Traditional Banking Context in Pakistan

With guidance from our industry mentors, we defined the control group to reflect Pakistan's conventional credit assessment approach. While traditional banks like Meezan typically rely on the State Bank of Pakistan's Electronic Credit Information Bureau (ECIB) data, we took a different approach. Instead of using the ECIB system that Meezan was initially supposed to provide, we utilized data that Meezan Bank had approved from HIES, which allowed us to maintain the authenticity of real-world banking data while ensuring proper data governance. The traditional ECIB-based system is notoriously sluggish: it can take several days to process just ten loan applications, since each file must be manually reviewed. Ultimately, approvals depend heavily on the individual officer's judgment and experience. As a result, even if a bank aims to approve 500 loans, different officers may reach different conclusions simply to safeguard against risk. This inconsistency leads to missed revenue opportunities—turning away creditworthy applicants means fewer loans are issued, which hurts the bank's bottom line in the long run.

Control Group Dataset Construction

For our control group dataset, we mirrored the standard bank practice of using simple yes/no flags to gauge creditworthiness rather than detailed risk factors. The dataset was sourced from HIES data that had been approved by Meezan Bank, ensuring authenticity while maintaining data privacy standards. This approach allowed us to access comprehensive Pakistani demographic and financial characteristics without compromising personal information, while enabling our model to identify meaningful patterns.

Data Processing and Mapping

The original dataset required extensive preprocessing and mapping. We received data with columns that didn't directly correspond to our required features, necessitating one-to-one mapping using unique identifiers to create a comprehensive CSV file suitable for analysis.

Dataset Scope and Filtering

- **Initial Dataset Size:** Approximately 160,000 records
- **Filtered Dataset Size:** 93,000 records (adult population only)
- **Filtering Rationale:** Banking industry credit scoring primarily targets adults who are eligible for bank accounts and loan products. Since credit assessment fundamentally involves lending to account holders who meet eligibility criteria, we focused exclusively on adult demographic data to ensure relevance to real-world banking applications.

This filtering approach aligns with industry standards where credit scoring models are designed specifically for the adult population who can legally enter into financial contracts and maintain banking relationships.

5.2 Experimental Group:

The experimental group incorporated comprehensive data from customers, including demographic details (Age, Gender, Marital Status), residential and occupational information (Residence Status, Employment Type, Industry, Designation), financial specifics (Gross Monthly Income, Monthly Expenses, Asset Value, Required Financing), and historical banking activity (Account tenure, Loan history). Data columns explicitly leveraged included:

- Numerical attributes such as Monthly Rent, Monthly Income, Net Savings, and Loans taken and repaid.
- Categorical features including Education Level, Workplace sector, and Reference Relationship.

This enriched data set enabled detailed analyses and precise credit score predictions leveraging advanced machine learning algorithms.

5.3 Project Evolution

A critical dimension of this project's evolution was the shift in scope and service orientation — from a general loan credit scoring system toward a specialized, targeted solution for Meezan Bank's **MyBykea** loan service, a recently introduced product designed for a specific customer segment. This pivot was not without challenges:

- **Data Scarcity and Integration Challenges:** Transitioning to MyBykea meant dealing with sparse or incomplete datasets, as this service was new and lacked extensive historical data. Our team had to innovate around data collection, cleansing, and augmentation methods to build a usable dataset.
- **Management and Strategic Direction Shifts:** The bank's management, realizing the need for a more focused, nimble solution, actively influenced the project scope. This entailed frequent strategy meetings and realignments of project milestones to meet evolving business priorities.
- **Architectural Overhaul:** Initially, our system was architected as separate microservices, designed for modularity and scalability. However, upon the recommendation of Meezan Bank's leadership, we pivoted to a monolithic architecture. This shift was motivated by:
  - **Streamlined Operations:** A monolith simplified deployment pipelines and reduced system complexity, crucial given the tight timelines and changing requirements.
  - **Improved Stability and Maintainability:** The monolithic approach allowed for easier debugging, tighter integration between components, and faster iteration cycles, which were essential as the project adapted to new data and business constraints.
- **Software Engineering Adaptations:** This architectural shift required significant re-engineering efforts, including codebase restructuring, database schema revisions (with PostgreSQL as the primary data store), and revamped API designs. These adjustments, while challenging, ultimately enhanced the robustness and responsiveness of the deployed solution.

This entire process showcased our team's agility and technical expertise, managing real-world complexities and shifting priorities while delivering a viable, high-performance AI credit scoring service tailored for Meezan Bank's innovative MyBykea product.

5.4 Software and Hardware:

Software Environment:

Backend Framework:

We began prototyping with a microservices-style Django setup—separating authentication, API endpoints, and business logic into independent services. This modular approach allowed parallel development, but it quickly proved cumbersome: coordinating multiple repositories, deployment scripts, and inter-service communication added overhead.

To accelerate development and simplify integration, we then consolidated everything into a single, monolithic Django application. This "all-in-one" architecture let us iterate faster—changes to models, APIs, or business logic could be made in one place without worrying about version mismatches between services. After a few sprints, however, Meezan Bank expressed interest in a more "professional" microservices approach for long-term scalability. We prepared a plan to isolate user-facing APIs, credit-scoring logic, and data-access layers into separate services, each with its own Docker container and CI/CD pipeline.

Ultimately, because Meezan's team valued rapid iteration and ease of maintenance over a fully containerized ecosystem, we remained on the monolith. This satisfied their requirement for quick turnaround on feature requests, while still leaving the door open to refactoring into microservices later if needed.

- **Database Management:**
- **SQLite (Initial Prototyping):** During the early monolithic phase, we used SQLite as a lightweight, file-based database. Its minimalist setup let us modify schemas and test API endpoints without provisioning a database server.
- **PostgreSQL (Production-Ready):** As the volume of data grew and performance requirements solidified, we migrated to PostgreSQL. This transition involved updating Django's DATABASES setting, adjusting data types and writing migration scripts to port existing SQLite data into PostgreSQL. With PostgreSQL's advanced indexing, transactional integrity, and support for concurrent connections, the application could handle real banking data securely and efficiently..
- **Frontend Technologies:** Tailwind CSS powered our responsive UI, enabling rapid prototyping of loan-application forms, dashboards, and approval workflows. We combined

Tailwind with Django's templating engine for server-rendered pages, then progressively introduced minimal JavaScript for form validation and dynamic dropdowns.

- **Machine Learning Libraries:**
- **MLPRegressor Findings:**
  - After splitting the cleaned dataset into 80 percent training and 20 percent test sets, the three-layer MLP (128 $\rightarrow$ 64 $\rightarrow$ 32) converged in under 200 epochs. On the test partition, it achieved:
  - Mean Squared Error (MSE): 24.0888
  - R-squared (R^2): 0.9780
  - Other layer configurations were also trialed—(256, 128), (64, 32, 16), etc.—but (128, 64, 32) yielded the best balance of accuracy and training time.
- **HistGradientBoostingRegressor Findings:**
  - We combined multiple subsets of historical application data (e.g., dfa and dfb) into a single training frame, applied the same preprocessing steps, and trained a HistGradientBoostingRegressor with default parameters plus sample weighting. On a dedicated validation split, this model produced:
  - Mean Squared Error (MSE): 3.1200
  - R-squared (R^2): 0.9971
  - Intermediate runs showed steady improvement as we tuned the learning rate and tree parameters (e.g., reducing MSE from 15.67 $\rightarrow$ 6.46 $\rightarrow$ 5.42 $\rightarrow$ 4.97), culminating in the final 3.1200/0.9971 result. The high R^2 indicates that the gradient-boosted tree was nearly perfect at capturing nonlinear interactions in the data.
- **Deployment Platforms:** Potential deployment discussed includes AWS, Azure, or Meezan Bank's proprietary infrastructure.

Hardware Infrastructure:

- Cloud deployment via AWS or Azure, leveraging scalability, availability, and robustness. Could be potentially used however below point clarifies!
- Internal deployment to Meezan Bank's data center infrastructure remains a strategic consideration, providing enhanced data security and control.

5.5 Parameter Settings:

Optimal hyperparameter tuning was executed systematically to achieve maximum predictive accuracy, the following results are based on several hyperparameters that we tested and fine tuned for our model.

Model Type	Configuration	Data Size	R^2 Score	MSE
MLP Regressor	hidden_layer_sizes=(256, 128), max_iter=200, random_state=42	30,000	0.9709	31.8890
	hidden_layer_sizes=(128, 64, 32, 16), max_iter=200, random_state=42		0.9779	24.1515
	hidden_layer_sizes=(128, 64, 32), max_iter=200, random_state=42		0.9780	24.0888
	hidden_layer_sizes=(64, 32, 16), max_iter=200, random_state=42		0.9856	15.7537
	hidden_layer_sizes=(32, 16), max_iter=200, random_state=42		0.9731	29.3963
	hidden_layer_sizes=(64, 32), max_iter=200, random_state=42		0.9875	13.7103
	hidden_layer_sizes=(128, 64), max_iter=300, random_state=42		0.9901	10.8477
HistGradientBoosting	learning_rate=0.08, max_iter=2000, max_depth=7, min_samples_leaf=10, l2_regularization=0.1, early_stopping=True, random_state=42	80,000	0.9854	15.6680
			0.9940	6.4635
			0.9944	5.9858
	learning_rate=0.08, max_iter=2000, max_depth=5, min_samples_leaf=10, l2_regularization=0.1, early_stopping=True, random_state=42		0.9950	5.4214
			0.9954	4.9688
Validation (Final Model)	Best HGBoost configuration	13,000	0.9971	3.1200

Help Screen

These comprehensive settings were determined via exhaustive grid-search validation, enhancing the model's predictive fidelity.

5.6 Data Preprocessing:

Our preprocessing pipeline was rigorously structured, involving multiple systematic stages:

- **Imputation:**
  - **Numerical Data:** Mean-value substitution for missing data, ensuring stability in continuous variables.

    - **Categorical Data:** Most-frequent-category substitution, preserving data distribution consistency.

- **Data Scaling and Normalization:** StandardScaler employed for numerical attributes, standardizing the data to ensure model stability and accurate predictive analysis.

- **Categorical Variable Encoding:** One-hot encoding implemented via OneHotEncoder, effectively transforming categorical features into machine-readable numeric formats. testing.

- **Data Cleaning:** Custom scripting rectified inconsistencies in attributes such as Residence Status, Education Level, and workplace sector, ensuring data integrity and accuracy.

These thorough preprocessing efforts ensured our model's predictive accuracy and generalization across diverse financial profiles.

5.7 Evaluation Metrics:

To validate our model comprehensively, multiple statistical metrics were employed:

- **Mean Squared Error (MSE):** Directly measured predictive accuracy by penalizing larger prediction deviations heavily.

- **R-squared (R^2):** Explained variance, quantifying the proportion of credit score variability explained by the model analysis.

Cross-validation methods were employed extensively to prevent model overfitting, ensuring reliable performance across multiple data subsets.

5.8 Experimental Setup:

The methodological flow involved structured and methodical processes:

- **Data Gathering:** We worked with a fully private dataset of around 93,000 customer records containing financial and loan information. To protect privacy, all personal identifiers were anonymized and mapped internally by our team—no sensitive personal information was exposed at any point.

- **Thorough Preprocessing:** We carefully handled missing data, scaled numerical features, encoded categorical variables, and cleaned inconsistencies to ensure the dataset was accurate and reliable.

- **Model Training:** Using this processed data, we trained neural network models (MLPRegressor) tuned to capture complex credit risk patterns.
- **Evaluation and Benchmarking:** Model performance was validated on a reserved set of 13,000 records. Our AI predictions were rigorously compared against Meezan Bank's traditional credit scoring methods, demonstrating significant improvements in accuracy and consistency.

Chapter 6

Results and Discussion

Our comprehensive evaluation highlights the success of our innovative semi-supervised ensemble approach, which blends teacher-student learning with weighted HistGradientBoosting to revolutionize credit scoring. The system was rigorously tested across multiple configurations, using three distinct data splits (A and B for training, C for validation) to ensure robustness. Below is a summary of our key findings:

Model Configuration	Hyperparameters	MSE	R^2 Score
MLP Regressor (Baseline)	layers=[64,32], dropout=0.2	10.521	0.9823
Initial HGBoost	random_state=42	15.6680	0.9854
Tuned HGBoost	max_depth=7, l2_reg=0.1	6.4635	0.9940
Weighted HGBoost (Final)	learning_rate=0.05, max_iter=2000	4.9688	0.9954

The teacher-student framework emerged as a cornerstone of our methodology, offering three distinct advantages:

- **Domain Expertise Integration:** The scoring formula captured critical financial heuristics, ensuring our model was grounded in real-world credit assessment principles.
- **High-Quality Training Data:** By generating reliable initial labels, the formula provided a solid foundation for the machine learning model to build upon.
- **Seamless Hybrid Approach:** The framework bridged the gap between rule-based systems and data-driven models, yielding a solution that was both interpretable and powerful.

6.1 Implementation Highlights

To ensure optimal performance, we processed 32 features spanning financial, demographic, and employment data. Key steps included:

- Standardizing numeric features (e.g., income, savings, and debt amounts) to maintain consistent scales.
- One-hot encoding categorical variables (e.g., education level, employment type) to capture nuanced relationships.
- • Applying differential weights (1.0 for formula-labeled data, 0.2 for model-predicted data) to prioritize high-confidence samples during training.

6.2 Validation Outcomes

The final model demonstrated exceptional reliability across diverse scenarios:

- **Consistency:** Delivered uniform performance for applicants of varying demographics.
- **Robustness:** Handled edge cases gracefully, such as applicants with no reported income or exceptionally high savings.
- **Speed:** Achieved sub-2-second prediction times, making it ideal for real-time credit assessments.

6.3 Model discussion based on experimentation

The Why Factor:

This part will be more focused on the model related discussions. Starting off with the why, i.e. why go for a neural network approach as a start. Upon recent experimentation and research work, we came across several solutions that favored neural network approaches as the standard for all kinds of AI/ML solutions. Particularly with regressor models, the benefits Neural Networks provide in terms of their deep learning abilities and capturing trends and patterns across layers and layers of hidden states, the approach allows for a way deeper understanding and tackling of the data related issues. While we did experiment with decision trees and a k-means approach in the beginning, we soon came to realize that the two were fairly shallow for the number of features and the amount of varying data trends we wish to envision and proceed with. Within neural networks we then decided upon MLP regressor which is basically a multi-layer neural network that enables even further understanding and decision making with regards to the data.

Next comes the Histogram Gradient boosting regressor. As we mentioned before neural networks were ideal for learning deep data trends that shallower models like gradient boosting might miss (shallow in comparison to neural networks, our ensemble approach however led us to the implementation of HistGradientBoosting as a key feature it enabled us to implement was weights. The issue with neural networks is that as it's a black box model, you aren't wary of the weights you assign as the model itself decides what's best, while this works, it prevents us from implementing our ideology. I will explain the experimentation logic ahead, but in simpler terms, it was HistGradientBoosting that allowed us to assign a higher weight and thus preference to the formula scored dataset as opposed to the model generated data.

Details perhaps missed:

To explain these, I will reiterate the logical approach of our model, to justify and clarify any confusions that may still linger.

With a curated unlabeled dataset of 93k records we split the data into three parts, let's label them A, B and C for now with the rows divided as A (50k), B (30k), and finally C(13k). We first scored A with our own credit scoring formula, this basically assigned weights to all the 30+ factors that we analyze in our app's form, the weights were assigned via logic and research, i.e. highest priority given to income, debt, expenditure etc. Once scored we then trained and fine tuned our MLP regressor model on this data until we received a reasonable threshold for our MSE and r-squared scores.

We then used this model to score B, once scored, we then concatenated our now labeled datasets A and B based on weights, 1 for A, 0.2 for B, the reason here being that, if we had firstly simply just trained on model on all of the 93k data, we would basically be replicating our formula meaning no real reason to even have an AI model, secondly, had we even done that, our model would that have to be retrained on new model generated scores and inputs in the near future to move away from the formula. Therefore, we implemented a student teacher distillation system, the teacher here would be the formula, that the student (MLP regressor model) would be trained to mimic, this would then be further distilled into our HistGradientBoosting model which would capture both formula-based scores and model-based scores, however give priority (weight 1) to the formula (our baseline) as that's what sets a standard for our scoring. Finally, this new model was then fine tuned and refined to reach an MSE score that's less than 1 percent of our possible score as well as a 0.99 r-squared score.

After this we then used C for validation, we ran the unlabeled C to be scored by both the formula and our final model, and then checked the MSE and r-squared on this as well. This again gave a really low MSE, lower than before even and a r-squared of 0.99, this signified that our model generalizes enough to score like our formula, but doesn't fully replicate the formula exactly therefore adding to its value as an AI model.

Further investigation:

We could definitely mess around with the data splits for starters, i.e. test different splits like 40-30-23k and so on. We could also try different models accordingly to achieve maybe an even better MSE and r-squared score.

The main thing that we could improve on is either having an actual certified scoring formula, labeled dataset, or at the very least financial data from 2023 onwards as our current data is 2018/2019 only, all of these could definitely ensure our model grows even further.

6.4 Limitations

While our approach achieved notable success, we recognize several areas for future refinement:

- **Data Constraints:** The formula-derived labels, though effective, may not encompass all nuanced risk factors.
- **Computational Demand:** Training required 2,000 iterations for convergence, highlighting a trade-off between performance and resource use.
- **Feature Dependence:** Accuracy depends on the precision of financial data reported by applicants.
- **Demographic Representation:** The training dataset may not fully capture the diversity of Pakistan's population.

Chapter 7

Conclusions , Limitations and Future Work

This section will detail the several issues we faced across the development of this project, issues we still have not overcome, and the future work that shall be required to eventually have our product reach a stage in development that makes it a complete and sellable product.

7.1 Limitations and issues faced:

Project genesis and scope:

The idea itself although proposed by Meezan Bank was too vague and could not be understood as a concrete answer for what was required, the base description stated an AI based credit scoring system, no further specifications were mentioned regarding the features, intended users, expected model behavior, or evaluation criteria. However, our team's interest in further pursuing the machine learning field piqued our interest in this project as this would also be a worthy learning experience to understand the functionality of a credible banking system.

After several meetings we came to an understanding that Meezan's requirements were simply a credit scoring system, the rest was up to us. To improve our project scope, we then integrated a dashboard system, a user-friendly UI which could be used by both Meezan employees and regular customers alike. We also sought to integrate a reporting system via a rag-based approach (will be discussed later on).

On the other hand, the project's trajectory was disrupted multiple times due to inconsistencies in stakeholder requirements as in a later meeting with Meezan Bank our scope was narrowed to a credit scoring system for their new bike loans program where we would target the credit scores of those specific applicants. We perceived this to be our final focus however in another meeting down the line we were instead told to dictate our app based on car loans. These abrupt changes created confusion, affected our development schedule, and hindered our ability to build a stable, well-scoped product.

Data (Major issue):

A massive need for any machine learning project is the data, this however ended up becoming our biggest problem in the end.

In our situation, we could have procured an online dataset at any point and trained our model on that, however that would then simply be a generalized credit scoring system. The need for Meezan bank's data specifically was to cater our application to Meezan bank's scoring model.

Our model which would be trained on the features and scores assigned by the bank would have been trained to think and act like a Meezan employee that would normally handle the scoring yet on a level above as it would also learn new underlying patterns across the data that even the employees miss, to generate scores that have an increased depth to their reasoning, what would function as a black-box model would eventually sway from the formula Meezan uses to purely focus on the existing record patterns to assign scores to new individuals.

Unfortunately, perhaps due to a communication or an understanding gap between our team and Meezan bank's team, the industry side refused to cooperate as they did not understand our need for their data. Across several meetings they told us to just generate fake data or to use a dataset online, failing to understand why exactly we need their data. At several points throughout the discourse of this project we asked for an AI/CS expert from their end to bridge the gap and translate our issues to them and theirs to ours, however this was never addressed and we were instead linked up with an individual from the Risk Assessment department who further complicated the situation due to a lack of understanding over AI model logic. Privacy concerns also arose as they worried our handling of the data may violate Shariah Compliance Laws. To address these, we offered them viable solutions, firstly offering to have our entire team sign an NDA to ensure the data wasn't misused, secondly, we asked for anonymized data meaning the records we would possess had no way of being linked back to any single individual as we didn't need the names, CNIC numbers or social security details of any individual. These records would contain data pertaining to statistics only, meaning including columns like income, asset value, residence status etc. With data as such that has no information relating to a certain individual, our data becomes fully anonymized as its now just financial data that could literally be anyone. Despite us addressing these issues, we were rejected any possibilities of the actual data for a while.

To cater for this we then acquired Meezan's bike loan form to get the relevant fields that customers might enter data for to apply for a bike loan, using this we generated logical synthetic data that we scored on our own devised scoring formula, that we then trained a model on, this however was also not accepted by Meezan as they didn't believe we had the expertise or educational background to accurately assign weights to score individuals. Due to the communication gap, they also believed a Blackbox model was an inherent negative which we couldn't convince them of to be false despite several discussions.

We then came up with three new possible solutions for Meezan. We prepared a HIES dataset collected by the Pakistani government containing details of household income statistics for houses and individuals for the years 2018/2019; by catering this to our needs we prepared a legitimate sample dataset that represented actual Pakistani individual's financial standings. This data was highly encrypted and had several data related issues and missing values etc. which had to be dealt with before proposing it as our actual dataset.

The proposed solutions included the following;

- Meezan bank gives us their scoring formula using which we can score our existing records and aim to match actual Meezan scoring logic.
- Meezan bank scores our HIES dataset and we use that to move ahead.
- We develop our own scoring formula through intensive research and credible sources and apply that on the dataset.

Once we have this prepared dataset, we can thus train our model on it accordingly, a possible issue raised here would be that the model would simply be scoring the newer records based on the formula itself, this would hold true to an extent but only at the beginning, as the model is trained on further model generated data, it would eventually converge to a point where its no longer just scoring based off a formula but instead based on hidden correlations between the data points to generate scores. Actual data from the bank wouldn't completely mitigate this but would ensure that the formula they used is by a professional expert in the field as opposed to us students, however although we did sign the NDA provided by Meezan Bank for dataset access, almost a week later, Meezan told us again to revert back to our own dataset for the model, they also didn't entertain the idea of scoring our data or approving our scoring formula for the data, leaving it up to us to decide how our model and formula decides a score.

This as we know threw a huge spanner in the works, with our minds fully committed to the possibility of actual Meezan Bank data being used to train our model, we were back at square one, and so as a collective and after the approval of our supervisor, we deemed the HIES 2018/2019 dataset collected by the Pakistani government to function as our eventual final dataset. This was incredibly challenging to decipher as the data had to be decrypted to make it plausible and understandable, secondly, we now had to retrain our whole previous model as well as our purely synthetic data approach raised concerns, these problems however led to innovation which is discussed later on, i.e. a semi supervised labeling approach with teacher student distillation as well as weighted ensemble training. The formula to score it initially however would be our own logically defined formula instead of Meezan's.

As mentioned before, had we been aware of the issues we'd face with financial data sharing, there was a high possibility we'd approach the problem with a completely different mindset and plan, something we will ensure to be well aware of when and if working with any banks in the future.

Model:

We now look towards issues with the model. Due to it being a relatively large amount of testing, we had to utilize free GPU access via Google Colab and Kaggle to further clean, enhance and train our dataset, access to a fulltime GPU could have greatly eased up this burden.

Coming to the actual model, we browsed through several possible ML algorithms i.e. KNN, decision trees, Linear regressors etc. and eventually ended up on Neural Networks which defers from the explainable AI side of things ensuring a Black-box model for scoring. Testing several hyperparameters on this was also relatively time consuming especially since we had to retrain our whole model on the dataset for different models and hyperparameters.

The major logical barrier we faced was the following, assuming we simply train our model on an MLP regressor, would it not simply be replicating the score we used? This would eventually stop being an issue as more and more scores generated by the model would in turn be used to feed the model itself, but who can wait that long, to get enough data to satisfy this need. To address this we came up with a new innovative idea, to firstly not score the whole dataset in one go, but instead create splits, this was discussed in previous sections.

Another issue now was what model to use for the second student part of the distillation process, after enough research we landed upon the Histogram gradient boosting regressor, this solution offered us weights. The problem is that with such models, if we give as much priority to the model generated data as the formula, we can risk overfitting, thus we assigned weights via this new model approach with a higher preference to the formula generated scores.

Webapp Backend:

For the Backend we used Django, as it offered a wide range of libraries (authentication modules etc.) and APIs to develop a webapp, getting to know and learn its functionality was a challenge due to our unfamiliarity with the tech stack.

Within Django, a major issue we encountered was connecting the backend logic to the frontend logic for which we struggled with what tech stack to implement, eventually deciding on HTML and Tailwind (a CSS library) for styling.

We also wished to integrate a RAG based model for the report generation however due to it using an open source resource intensive LLM (Qwen 2.5), it had to import to generate the report, our system's computational resources could not handle that workload, so we instead tried several other LLMs of variable sizes , some generated poor reports, others were again too heavy for processing on local systems so we ended up at a lightweight Gemini LLM called through an API key for report generation.

Webapp Frontend:

The frontend was ten times more challenging as opposed to the backend in this situation. For starters several libraries were very limited in terms of the functionalities they offered and what we were looking for; after browsing through multiple options, we chose HTML coupled with a CSS library named Tailwind. We aimed to recreate our proposed layout in the wireframes exactly for our frontend however handling containers etc. was a nuisance.

Our dashboard itself became a major issue as we couldn't figure out how to import charts onto our webapp, as well as integrate our charts to reflect statistics from both our existing data and as the user's inputs, for both of these we put a lot of effort into our views.py file, this was the gateway to connect our backend and frontend seamlessly as it used our backend logic to pass relevant context to our frontend html pages and caused several errors when configuring.

The reports page was also really difficult to integrate as our generated report through the API kept falling out of the container's range on the frontend, secondly, we wished to implement a feature to save and print your reports which also caused problems as the reports kept getting cut off when being saved or printed.

Another issue was that after our reversal to the HIES data, we had to navigate the charts to match that specific dataset's columns and integrate new and remove older charts as a whole.

DB integration:

By default, Django uses SQLite3 as a database which also offers a great admin panel which can be used to handle data and records from the admin side, however this was integrated into our main app itself and we wished to separate our DB and our main app.

For this we referred to PostgreSQL, this enabled us to have a separate PGadmin application to instead control and manage our data as well as also allowing us to perform SQL based queries for analysis on the said data, a feature SQLite3 wasn't offering, re integrating our entire data into this new DB caused several errors. We also tried importing our existing input form data from SQLite3 into our Postgres DB but weren't able to due to formatting issues.

7.2 Future Work:

While this project can be a solid product to be used by either Meezan employees or general customers, depending on who Meezan targets it to, there are still several improvements that could be made to the actual product to perfect it.

UI/UX:

For starters, while UI and UX had been cared for to a massive extent in general, consistency across pages, efficient error handling and warnings, a theme resembling Meezan's colors but also being different enough to be eye catchy and easy in terms of visual load. There were still changes we can add to make our UI reach its peak.

Possible additions:

- Implementing the multiple languages feature to cater for users that aren't familiar with English.
- Dark mode based on user preference.
- Option to make text and icons bigger for visually impaired individuals.
- Text to speech and speech to text capabilities for impaired individuals.
- Customizable layout i.e. color and design choices for users to make the app more interactive.
- Implementing a timezone feature..
- Implementing the profile visibility feature.

Data:

With the use of actual and increased Meezan Bank's financial data, our model would greatly improve as it could directly then act and think like a Meezan employee when dictating scores for individuals. Having reached an agreement on the data and having signed the NDA, it was unfortunate to be turned down however even at this stage if our model is now fine tuned on Meezan's data, although not fully, it could learn a few potential correlations and patterns that Meezan might unknowingly implement.

Training on actual data would also work amazingly as individuals who apply could be ensured that they won't face any bias for their scoring while also being judged as if by a Meezan employee thus getting an idea of their current standing in relation to other previous applicants who also got their scores assessed. They can thus see what they need to improve on to get a better score and what score they exactly need to pass the criteria for a bank/car loan.

By training our model on existing Meezan data, we could also guarantee ourselves a fail-safe for avoiding expertise concerns that the bank initially had, as the scoring would have already been done by their own employees, they wouldn't have to worry about us misassigning weights due to a lack of knowledge in the field.

As of now, our model is based off the HIES dataset however Meezan bank's data could turn thi generalized model to a more Meezan centric model catered to handling the bank's clients.

Report generation:

Our current reliance on the Gemini API is only a short-term solution as we keep needing to use new API keys when our usage limits expire, to counter this our Rag model works perfectly as it uses open source LLMs that can perform the same tasks for an unlimited period of time without complaints.

For this usage we need GPU access which If incorporated by Meezan Bank in the future could remove this liability on a whole ensuring a full centralized system with no need for free trial access keys for their report generation model.

7.3 Conclusion:

To end this report, we will discuss our final thoughts and experience with this project.

What once seemed to be a simple enough task was at times made unachievable due to industry related contractual issues. It was through our supervisor's guidance and our own willingness to learn and achieve something that we came to finalize this year long grueling experience. The project gave us great insights into how the industry functions, what problems you'll likely face when dealing with financial institutions and also how to overcome new learning curves and how to function as a team to achieve a significant goal in the end. It also gave us a great experience boost with designing fully functional web applications and an ai model that meets industry standards.

With this we conclude our report, thankyou for your time.

Chapter 8

References and Bibliography

HIES dataset URL

https://www.pbs.gov.pk/content/pslm-hies-2018-19-microdata

Meezank bank bike loan form

https://www.meezanbank.com/wp-content/themes/mbl/downloads/Apni-Bike-form.pdf

Other resources accessed

Analysis, Loan. "Commercial Loan Analysis: Principles and Techniques for Credit Analysis and Lenders (Bankline Publication) - Pirok, Kenneth R.: 9781557387165 - AbeBooks." Abebooks.com, 2025, www.abebooks.com/9781557387165/Commercial-Loan-Analysis-Principles-Techniques-1557387168/plp. Accessed 23 May 2025.

"Benchmarking State-of-The-Art Classification Algorithms for Credit Scoring on JSTOR." Jstor.org, 2025, www.jstor.org/stable/4101754, https://doi.org/10.2307/4101754. Accessed 23 May 2025.

Crook, Jonathan, et al. "Recent Developments in Consumer Credit Risk Assessments." European Journal of Operational Research, vol. 183, 2025, pp. 1447–1465, www.research.ed.ac.uk/en/publications/recent-developments-in-consumer-credit-risk-assessments. Accessed 23 May 2025.

"Credit Scoring and Its Applications, Second Edition." Google Books, 2017, books.google.mk/books?id=7UQzDwAAQBAJprintsec=copyrightv=onepageqf=false. Accessed 23 May 2025.