Convolve 3.0: A Pan IIT AI/ML Hackathon

Team

Summit Waves

Members:

Rahul Rajagopal & Raghav Chugh

Manipal University Jaipur - 2nd Year

IDFC FIRST Bank

A brief introduction and understanding of the problem

In the recent financial sector, effective management of credit card portfolios is among the greatest challenges for most banking institutions. The growing adoption of credit cards has also heightened the demand for effective risk control systems capable of forecasting customer defaults and providing remedial measures.

The system was developed using AI and is referred to as a Credit Card Behavior Risk Management System. The system aims to give a solution to financial institutions an all-in-one approach for determining whether to extend customer's credit. The system uses advanced machine learning models to provide insights into the incurred costs of delivering services to customers by calculating the chances of them defaulting. It uses the FinWell Score metric, which is a single score that makes it possible to combine behavioral, transactional, and external credit data for risk measurement.

The aim is to develop an idea that will arm all the banking institutions with a simple calculation, which in turn will assist them in making rational business growth decisions, better-formulating credit plans, and ensuring better stability by lowering their losses from such parasitic acts. This system will enable banks to avoid some of the gaps posed by credit risk management problems such as default prediction, data imbalance, and interpretability.

Understanding the Problem

Credit risk assessment has been done using credit scoring methods that are very simple and do not fully understand the customer. This characteristic of the credit scoring model causes banks to confound a customer who is likely to pay back the loan/credit as against one who is a potential defaulter. This creates distortions such as offering credit to customers who are a credit risk or turning away customers who have a low credit risk without having to do so. Both errors can reduce profits and lower customer satisfaction.

- **Default Probabilities Calibration:**There is a multitude of such models, however a critical factor that most of them possess is the lack of the precise assignment of the default probability for customers to their respective categories due to the simplistic model design or the lack of integration of diverse behavioral attributes.
- **Comprehensive Risk Metric Development:**The problem is that such simple approaches of embedding the development of metrics of the likelihood of default are not sufficient. The banks require a metric that integrates multiple features such as on-us radius transactions, bureau aggregate insights, and inquiry aggregate behaviors including default probabilities. This has subsequently led to the invention of the FinWell Score which embodies all these aspects into a more coherent single snapshot of credit health.
- **Addressing Data Imbalance (Vagueness in data attributes):**In credit card datasets, thьерсr has been a concern of having too few defaulters and too many defaulters making it hard to bias machine learning models without hindering their accuracies.
- **Actionable Risk Categorization:**Customers must be listed in effective risk scales that can help in operational types of decisions (for example, Low, Moderate, High) which can then aid them in taking measures relevant to the target in question.
- **Scalability and Real-Time Performance:**Institutions such as these need to have solutions that can handle hundreds of millions of records and provide insights in close to real-time for fast decisions to be made.

The core issue being addressed and evaluated is risk management concerning credit card use. The main objective is to devise and train an AI predictive model designed to predict the risk of debt default based on customer spending behavior.

Our solution and its purpose

Through the utilization of Random Forest Algorithms, overreliance on manual work is eliminated as multidimensional feature scaling is incorporated through our machine learning model into the **FinWell Score** Predictor application which can assess the credit scoring of customers with ease and low risk. The customers are provided a score that determines their financial performance and based on that score, IDFC bank will have an appraisal of their credit risk, greatly helping them in their risk management.

The main purpose was to create a predictive model that calculates the probability of customers defaulting on their credit card payments. We realized a scoring metric or a range that can be personalized for the bank that the customers can view would be beneficial for both the bank and the customers. The bank would also understand its customers better for risk management which allows them to take appropriate actions on their existing credit card users as well as set parameters and better set eligibility norms for future credit card users. Although getting the defaulting probabilities of customers is useful, it can make it difficult to understand and process them for risk management at a glance for bankers. So we have come up with our own scoring metric and system called **FinWell score** that utilizes the probability and the other 4 attributes such as onus, transaction, and bureau attributes to have a range between 300 to 900. As you can see, the scoring metric is very similar to our usual credit score for easier understanding.

The FinWell score is a unique metric that can be tailored even further for the bank and has huge scaling potential. It is helpful for customers as they can keep a score of their financial discipline which can also open doors to various benefits. We have also built a functional user interface that the bank employees can use to obtain the FinWell score along with the probabilities in a CSV downloadable format. This ensures a fast & smooth way to conduct risk management analysis for their credit card customers.

- ➢ Developed a new unique scoring metric known as "FinWell Score" that is utilized for risk management - more features explained further in the documentation
- ➢ Designed and developed a working **user interface** for the risk management system that takes a customer CSV database and generates a new clean CSV with all customers and their associated risks along with their FinWell score.
- ➢ Also designed the IDFC bank mobile interface implementing the FinWell score using Figma.

Advantages for the Bank and Customers

Benefits for the Bank:

- **Proactive Risk Management**:
  - Early identification of potential defaulters reduces credit losses.
  - Enables targeted customer engagement and corrective measures.
- **Enhanced Profitability**:
  - Optimized credit limit assignments and interest rates improve portfolio performance.
- **Strategic Decision-Making**:
  - Data-driven insights empower better resource allocation and product design. The bank can confidently expand its lending portfolio by targeting low-risk customers.

Benefits for Customers

- **Personalized Financial Solutions**:
  - Risk-based credit offerings ensure fairness and transparency.
- **Improved Financial Awareness**:
  - FinWell Scores and risk categories help customers understand and improve their financial health.
- **Enhanced Services**:
  - Early interventions and tailored repayment plans provide support to at-risk customers. Customers can monitor and understand their FinWell Score to improve their financial health.

IDFC FIRST Bank

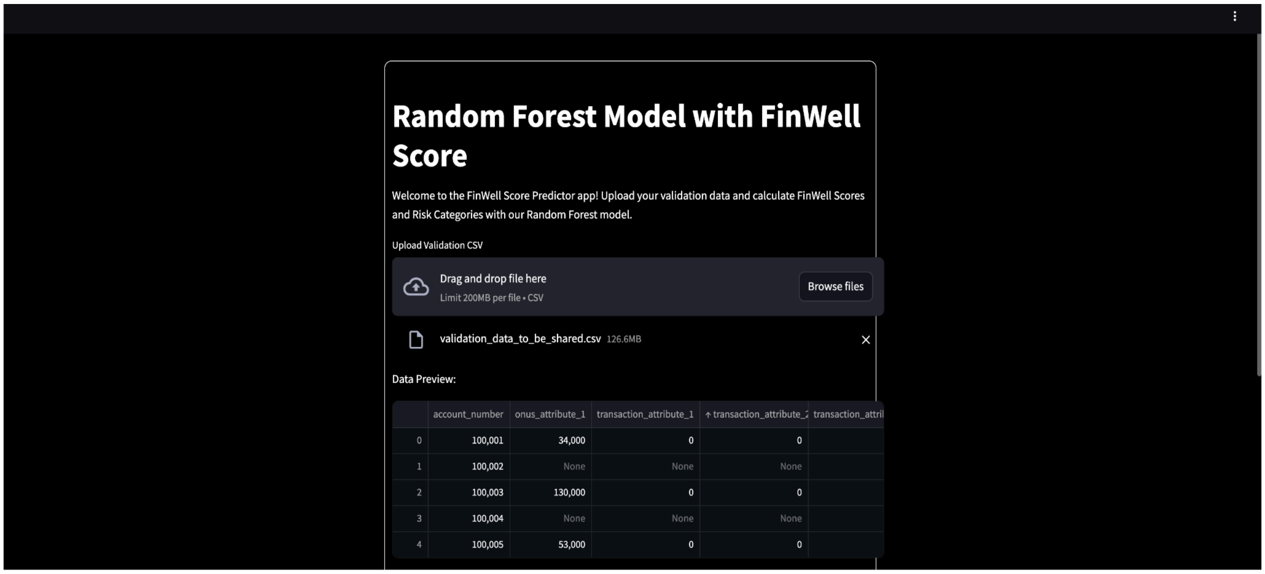
What have we worked on & developed?

- **Data understanding and pre-processing:**
    - Our first task was to understand the task at hand that was given to us and ideate on the specified problem statement. Once we did that, we focused on the two large data files that were given to us, one for creating the predictive model which contained an extra attribute named “bad_flag” and another file named “validation_data” that contained data on customers whose behavior score or defaulting probabilities we had to calculate using our model. This stage took a little while due to the sheer size of the data, the number of columns, and the overall vagueness of the complete data. Then, we focused on creating a mini roadmap or a task timeline based on the round 1 submission deadline which consisted of when we would complete developing the predictive model, other interesting ideas or insights that we could add & implement, make sure everything works and finally, the documentation.
- **Model Development process:**
    - We focused on analyzing the data and cleaning it to make sure there weren't any customers that had all null attributes or any errors in the data. This would ensure our model wouldn't encounter issues in parsing through the data. Then came choosing the right model that would calculate accurate defaulting probabilities based on their credit card activities (attributes in the data file) which was crucial. We tried a few different models such as logistic regression, random forest, xgboost (Gradient boosting models), and neural network (tensorflow) but we chose to utilize the random forest model. This gave great accuracy and worked as intended according to our plan, it was more accurate than logistic regression and more convenient to use than the later 2. We had run into issues when using xgboost and neural networks for the predictive model. The random forest model was also well suited for this type of data and its imbalance due to the sheer number of attributes. It can handle large datasets like these very efficiently resulting in great scalability potential. To summarize the libraries we used to develop with model, pandas, numpy, and scklearn were used.

- **Hyperparameter tuning for better accuracy:**
    - To make our predictive model more accurate, we decided to use hyperparameter tuning to obtain better accuracy and ROC-AUC scores. We needed to do this to understand and identify the best combination of parameters to improve the model. We decided to utilize GridsearchCV for our hyperparameter tuning because it was straightforward and convenient to implement and ensured a great performance boost based on the time constraints we had. We also tried improving the model using randomized searchCV but the process simply took way too long (upwards of 4 hours) which was too inefficient but also computationally resource-wasteful.
- **Development of FinWell Score formula and risk categories:**
    - As stated earlier in the documentation, we decided to come up with an interesting way for conducting risk management on the bank's existing credit card users called the FinWell score. This score combined the calculated defaulting probability and the other four attributes having different weights in a formula that would give values ranging from 300 to 900. We made this value and range similar to existing credit scores for easier understanding and adaptability. One of our main reasons for doing this was because the probabilities are not that easy & efficient way to conduct risk management and would result in an inconvenient method for bankers. So we decided to create the FinWell score and also categorize the risk based on the score for easy understanding. Initially, we used log transformation to come up with the FinWell score but that resulted in a very small range which was not that useful. Hence, we came up with a formula (which can be found in the code) that results in a value between 300 to 900.
- **Application Interface (Streamlit & Figma implementation):**
    - Built a user-friendly Streamlit app for seamless data upload, processing, and FinWell score calculating to produce a downloadable CSV file containing all customer account numbers along with their defaulting probability, FinWell score, and risk category. The interface is yet to be designed but this is just to showcase a working prototype. The implementation is shown below.
    - We also designed a simple mobile UI for IDFC bank showcasing the implementation of the FinWell score in their mobile app and its benefits.

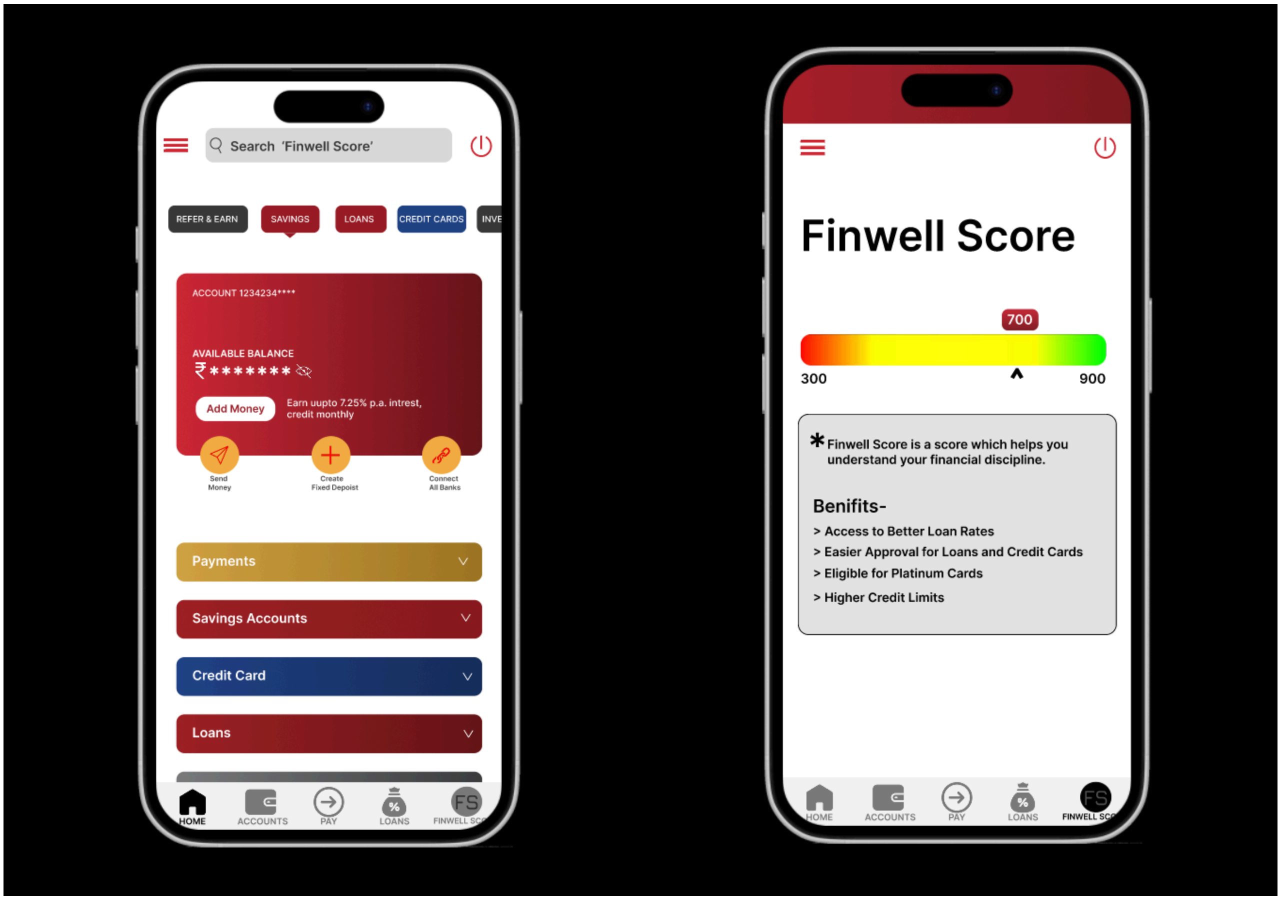
Streamlit implementation:

- This shows the streamlit implementation where the bank employee can upload a customer CSV and then it gets processed to generate a new clean CSV containing the account number, defaulting probabilities, FinWell score, and risk category.



IDFC FIRST Bank

Figma IDFC bank mobile UI implementation:



Additional Features and Insights

- **Real-Time Scoring**:
  - The system can process new customer data in real-time, enabling instant decision-making.
- **Customizable Risk Weighting**:
  - Banks can modify feature weights based on evolving business priorities and regulatory requirements.
- **Insights on the Financial Behavior of Customers**:
  - Generates actionable insights into customer behavior, enabling targeted marketing campaigns and product offerings. Customers with high credit usage and recent loan inquiries show a strong correlation with default risk.
- **Scalability**:
  - Easily scalable to handle large datasets and additional features, ensuring long-term applicability.
- **Transparency**:
  - Provides detailed breakdowns of scores and risk categorizations, fostering trust with customers and regulators. Maintaining multiple credit products with long histories of timely payments reduces risk significantly. Such customers demonstrate financial stability and responsibility.

IDFC FIRST Bank

Conclusion

The **Credit Card Behavior Risk Management System** is a transformative solution that has been designed to address key challenges in credit portfolio management. By using advanced machine learning techniques and introducing the **FinWell Score**, we have come up with a system that provides a scalable, efficient, and actionable framework for assessing customer credit risk.

For financial institutions, the system enables proactive risk mitigation, improved decision-making, and enhanced profitability. It helps banks to identify potential defaulters early, optimize credit policies, and build customer-centric offerings. For customers, the FinWell Score offers transparency, insights into financial health, and opportunities for improved financial management.

The approach that we have used, can integrate diverse data attributes and provide intuitive outputs, ensuring that banks can confidently manage credit risks in a dynamic financial environment. As a robust, real-time scoring system, it bridges gaps in traditional credit risk models.

Looking forward to hearing your thoughts!

Thank you!

Team Summit Waves