CAPSTONE PROJECT

CREDIT CARD DEFAULTER PREDICTION

Presented By:

Joshwa Thomas - Amal Jyothi College of Engineering - Computer Science and Engineering

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

Identifying potential credit card defaulters is essential for financial institutions to mitigate risk and manage credit effectively. This project focuses on developing a predictive machine learning model that leverages customer payment history and demographic information to forecast default probabilities. Accurate predictions enable proactive measures, improving credit management and reducing financial losses.

PROPOSED SOLUTION

- The proposed system aims to address the challenge of predicting credit card defaulters to minimize financial risks. This involves leveraging data analytics and machine learning techniques to accurately forecast the likelihood of defaults. The solution will consist of the following components:
- Data Collection:
  - Gather historical data on customer payments, including payment history, credit limit, and other relevant factors.
  - Utilize real-time data sources, such as changes in economic conditions and updates to credit policies, to enhance prediction accuracy.
- Data Preprocessing:
  - Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies.
  - Feature engineering to extract relevant features from the data that might impact the likelihood of default, such as credit utilization ratio and payment-to-balance ratio.
- Machine Learning Algorithm:
  - Implement a machine learning algorithm, such as a classification model (e.g., Logistic Regression, Random Forest, or Neural Networks), to predict the likelihood of default based on historical patterns.
  - Consider incorporating other factors like demographic information, account activity, and economic indicators to improve prediction accuracy.

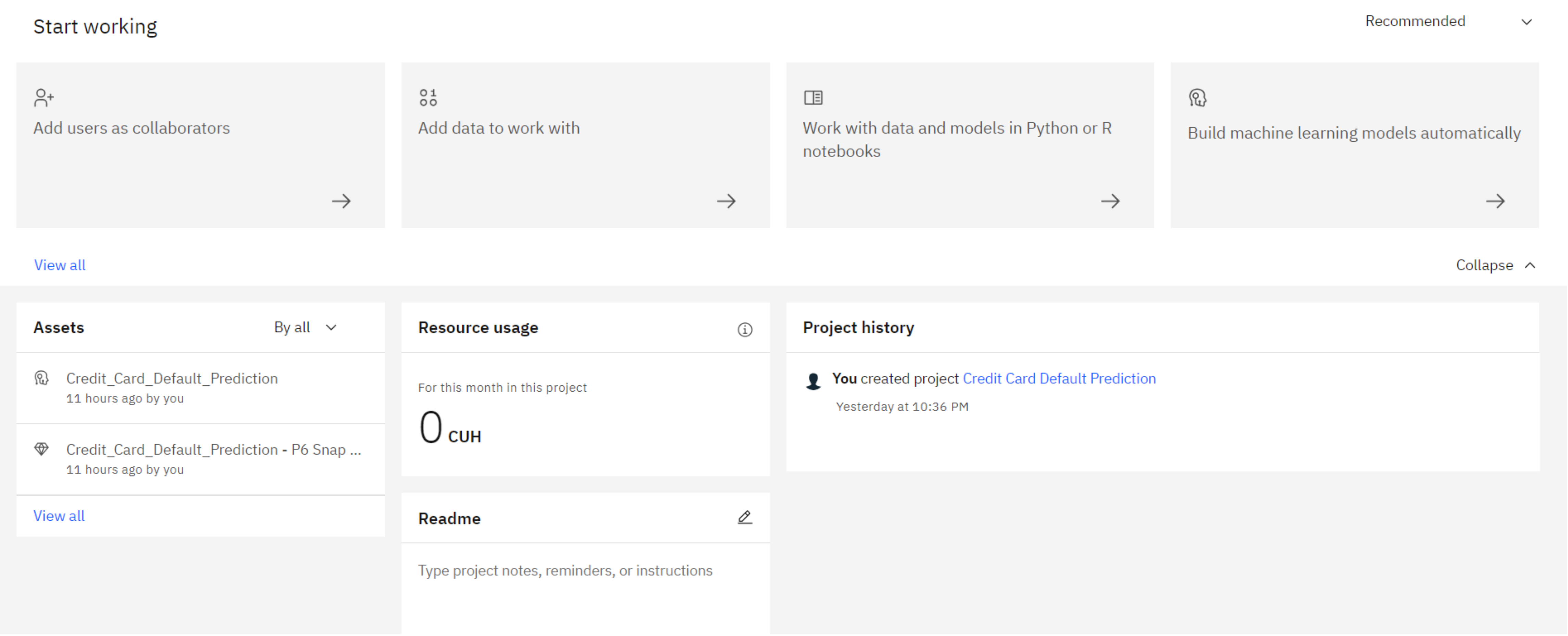
edunet
foundation

- **Deployment:**
  - **Develop a user-friendly interface or application that provides real-time predictions for default risk.**
  - **Deploy the solution on a scalable and reliable platform, considering factors like server infrastructure, response time, and user accessibility.**
- **Evaluation:**
  - **Assess the model's performance using appropriate metrics such as accuracy, precision, recall, F1-score, and AUC-ROC.**
  - **Fine-tune the model based on feedback and continuous monitoring of prediction accuracy.**

SYSTEM APPROACH

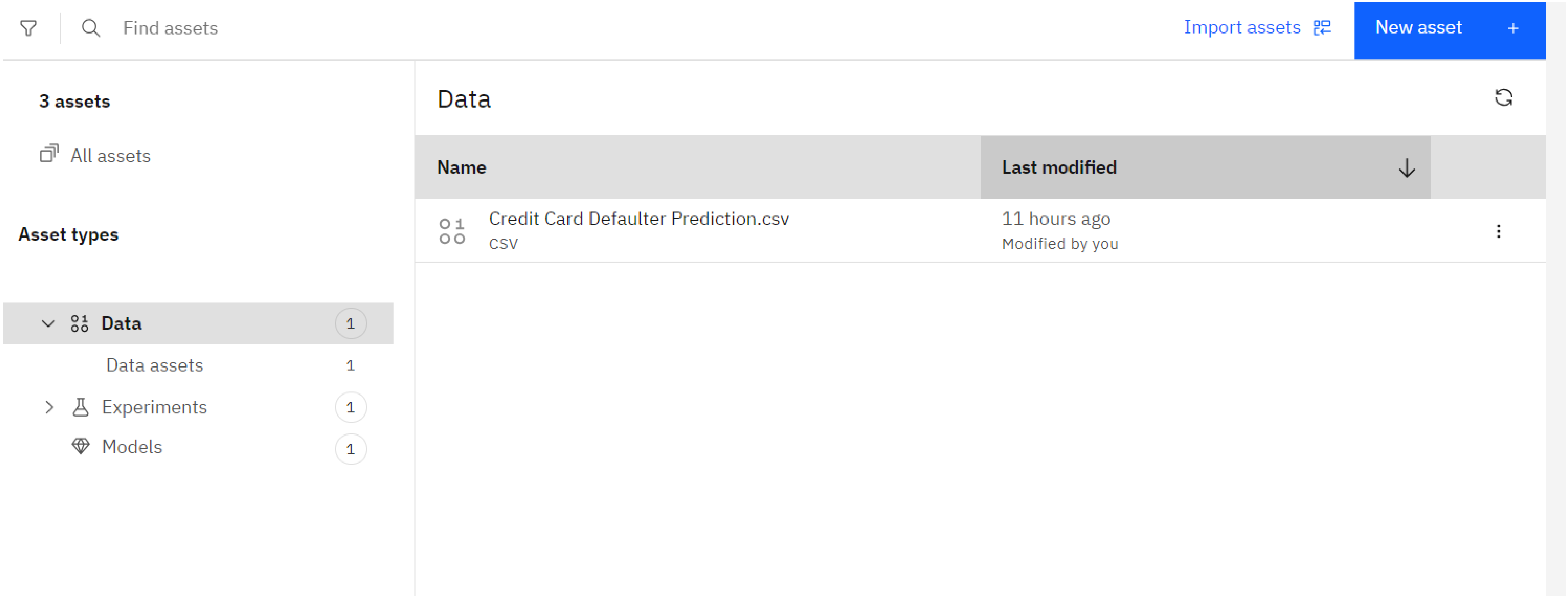
- **Data Collection:**
  - Historical credit card transaction data, including payment history, credit limits, demographic details, and previous defaults.
- **Technology:**
  - IBM Watson Studio AutoAI for automated data preprocessing, feature engineering, model training, and testing.
  - AutoAI's selection of algorithms and hyperparameter optimization for creating the most effective model.
- **Infrastructure:**
  - Cloud-based deployment using IBM Cloud for seamless scalability, reliability, and integration with IBM Watson services.

Working with IBM WATSON STUDIO



Data set uploaded into assets

A Credit Card Defaulter Prediction.csv dataset which contains fraudulent credit cards data is used as data assets here.



Model Experimenting Using AUTOAI

IBM WATSON MACHINE LEARNING using Auto AI will help in creating the credit card defaulter prediction model.

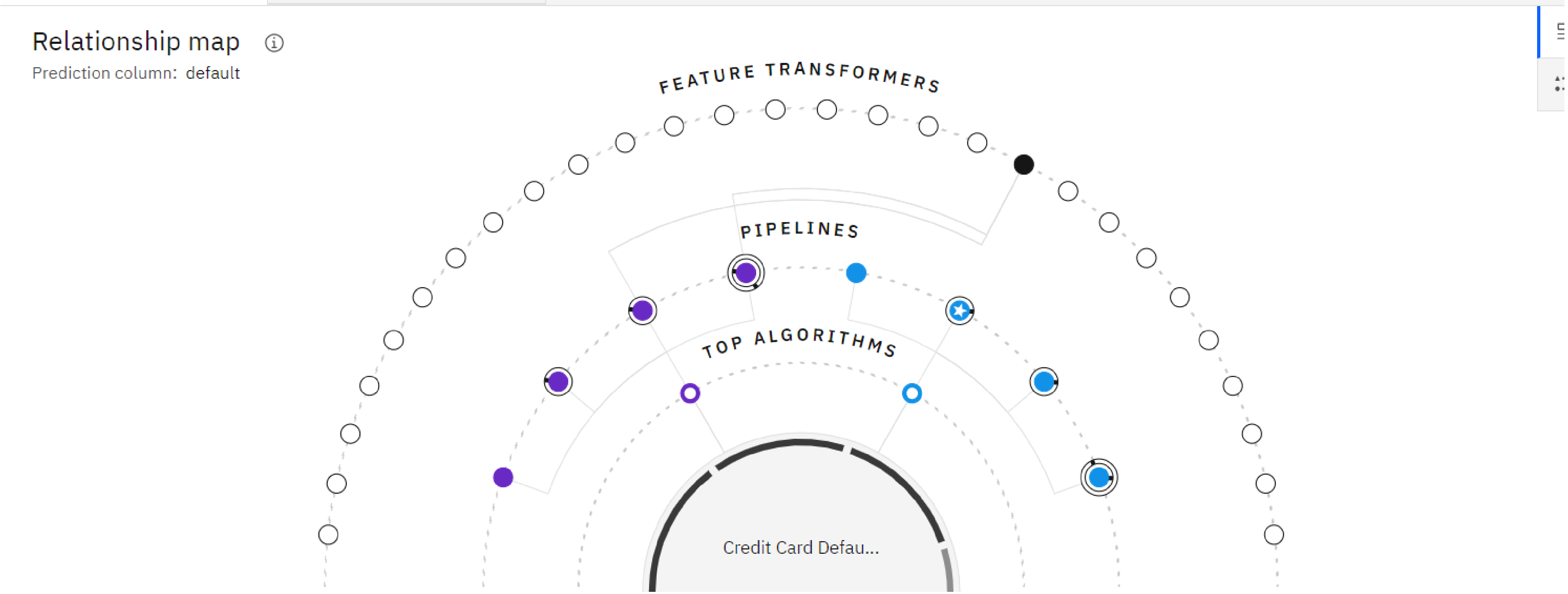
Experiments

Name	Status	Model type	Last modified
Credit_Card_Default_Prediction AutoAI experiment	Completed	Binary classification	11 hours ago Modified by you

ALGORITHM & DEPLOYMENT

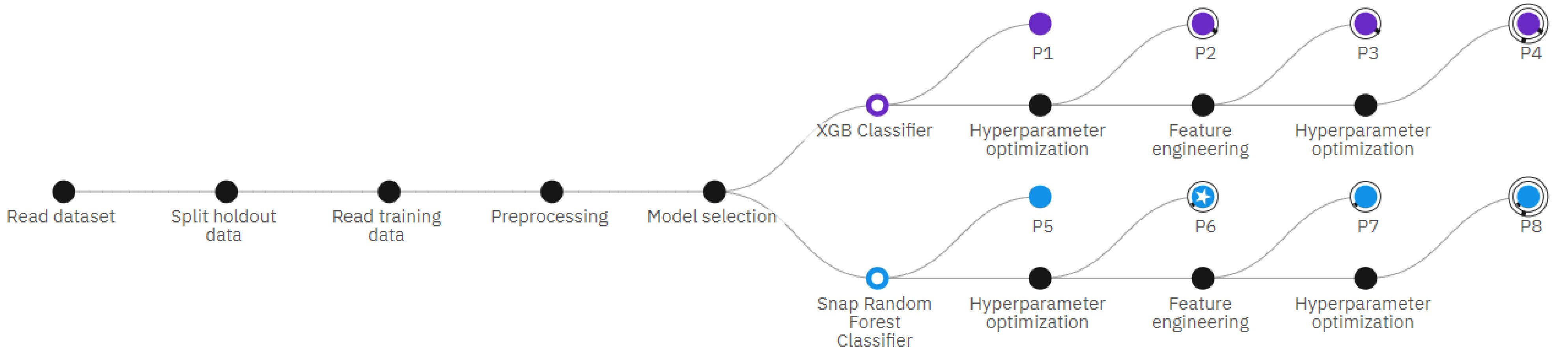
- In the Algorithm section, describe the machine learning algorithm chosen for predicting credit card defaulters . Here's an example structure for this section:
- **Algorithm Selection:**
  - A variety of algorithms were considered and evaluated using IBM Watson Studio Auto AI, with **Snap Random Forest Classifier** being selected as the final model due to its high accuracy and efficiency in handling imbalanced data. The choice was based on the model's ability to effectively distinguish between defaulters and non-defaulters using historical credit card transaction data.
- **Data Input:**
  - The model utilizes various input features, including credit limits, payment history, demographic information (such as age, sex, education, and marital status), past due payments, and billing amounts across multiple months.
- **Training Process:**
  - The model was trained using historical data on credit card transactions and customer information.
  - Auto AI facilitated automated feature engineering, hyperparameter tuning, and model selection, ensuring optimal performance.
  - Techniques such as cross-validation and stratified sampling were employed to handle class imbalance and evaluate model performance.
- **Prediction Process:**
  - the model predicts the likelihood of a customer defaulting on their credit card payment. Predictions are based on the input features and are updated as new data becomes available.
  - The model considers both historical and real-time data inputs, providing timely predictions for decision-making.
- **Deployment**
  - We use active space in deployments and make it online. We manipulate Json file .

EXPERIMENTING USING AUTOAI



Progress map

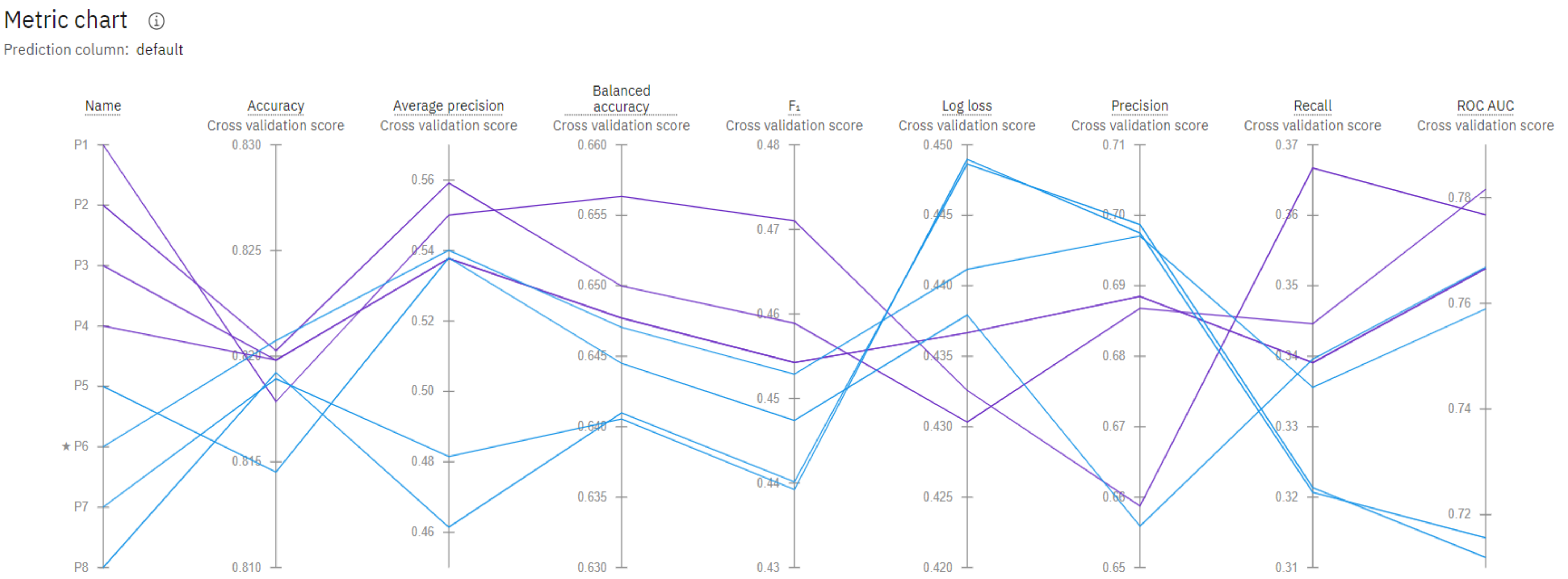
Prediction column: default



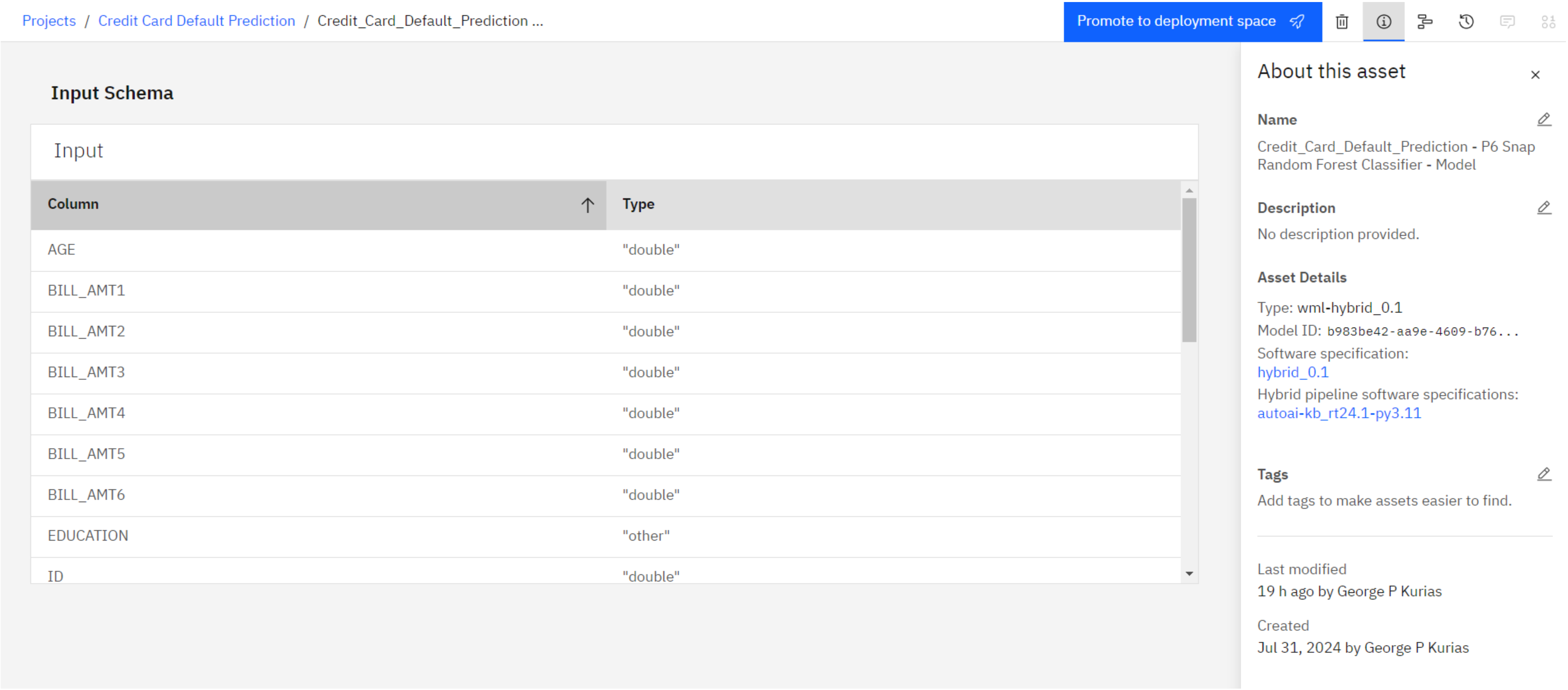
Pipelining rank : XGB classifier rank1

	Rank ↑	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	**Pipeline 6**	Snap Random Forest Classifier	**0.821**	HPO-1	00:00:23
	2	**Pipeline 2**	XGB Classifier	**0.820**	HPO-1	00:00:29
	3	**Pipeline 4**	XGB Classifier	**0.820**	HPO-1 FE HPO-2	00:02:07
	4	**Pipeline 3**	XGB Classifier	**0.820**	HPO-1 FE	00:01:37
	5	**Pipeline 8**	Snap Random Forest Classifier	**0.819**	HPO-1 FE HPO-2	00:02:02
	6	**Pipeline 7**	Snap Random Forest Classifier	**0.819**	HPO-1 FE	00:01:52
	7	**Pipeline 1**	XGB Classifier	**0.818**	*None*	00:00:04
	8	**Pipeline 5**	Snap Random Forest Classifier	**0.815**	*None*	00:00:03

Pipelining comparison



Deployment using active space



RESULT

- **Model Performance:**
  - Accuracy: 0.826
  - Precision: 0.706
  - Recall: 0.364
  - F1 Score: 0.481
- **Visualizations:**
  - Confusion matrix
  - ROC curve

JSON CODE :VALUES CHANGED FOR TESTING

{"input_data":

[{"fields":["ID","LIMIT_BAL","SEX","EDUCATION","MARRIAGE","AGE","PAY_0","PAY_2","PAY_3","PAY_4","PAY_5",

"PAY_6","BILL_AMT1","BILL_AMT2","BILL_AMT3","BILL_AMT4","BILL_AMT5","BILL_AMT6","PAY_AMT1","PAY_A

MT2","PAY_AMT3","PAY_AMT4","PAY_AMT5","PAY_AMT6"],"values":

//positive case input value

[[7391,20000,"M","University","Single",28,2,2,0,0,0,2,42670,42889,42689,42689,42689,42670,0,0,0,0,0,0],

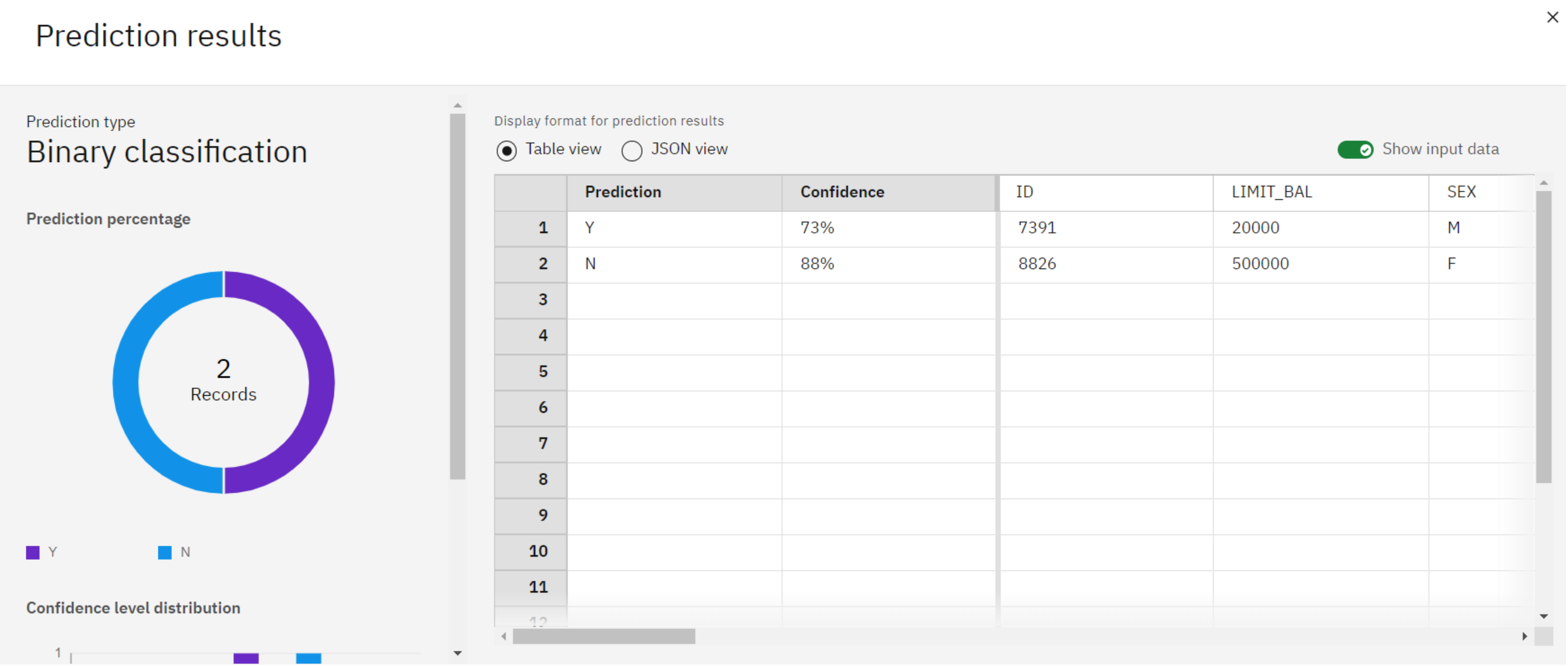
//negative case input value

[8826,500000,"F","GraduateSchool","Married",43,0,0,0,0,0,0,24751,28063,21388,21725,20760,19680,2221,278

0,3187,2232,2214,2672]

]}]

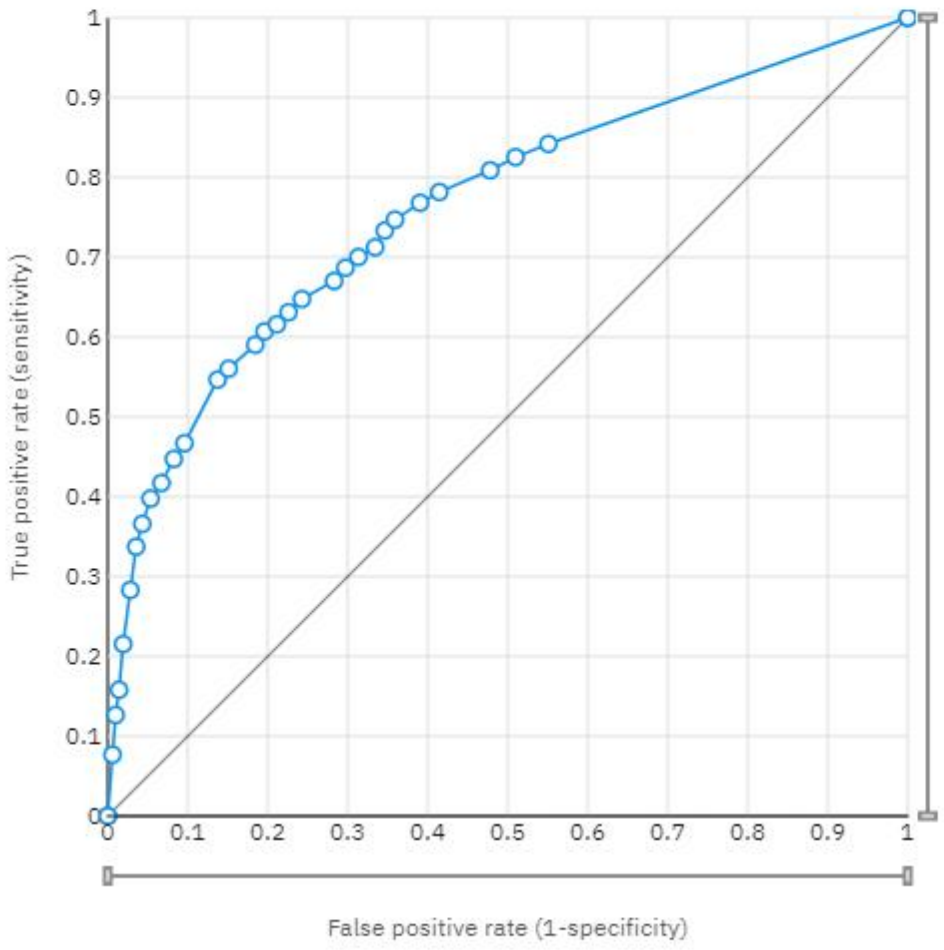
}



Y is to tell positive and N is to tell negative

Here , We can see that the model has correctly predicted the Output

ROC Curve



True positive rate (sensitivity)
False positive rate (1-specificity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

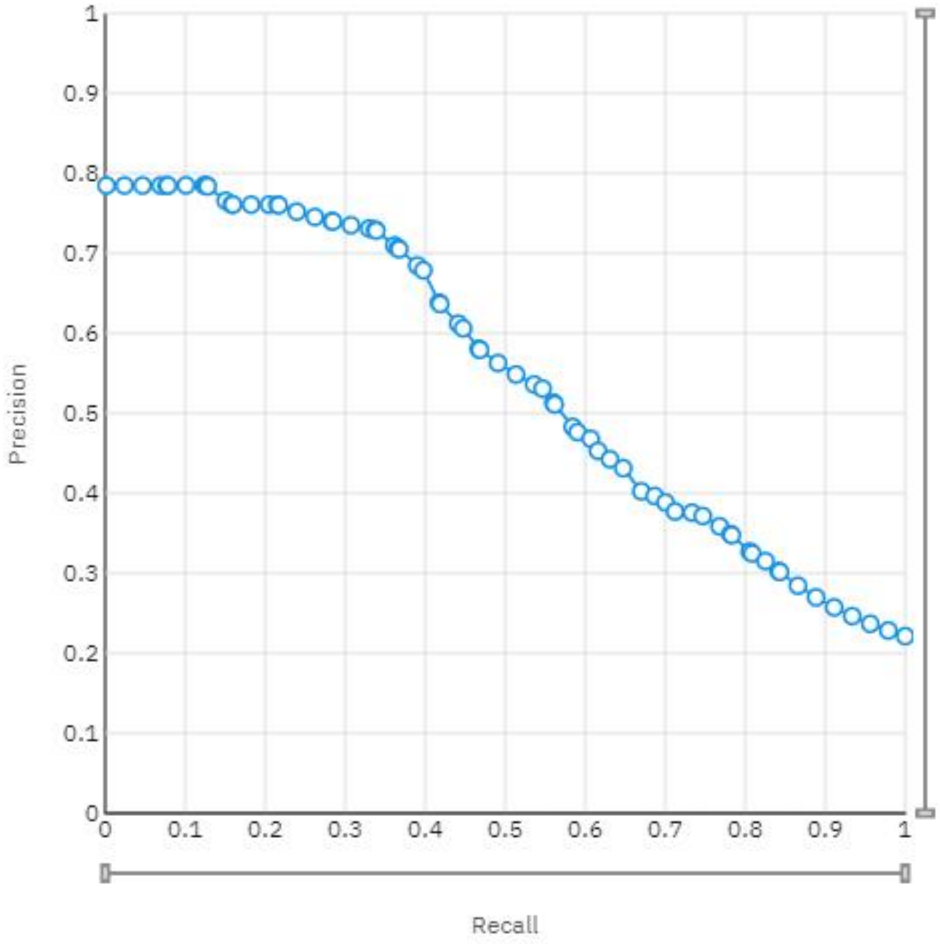
Confusion Matrix

Confusion matrix ⓘ

Observed	Predicted		
	Y	N	Percent correct
Y	242	422	36.4%
N	101	2235	95.7%
Percent correct	70.6%	84.1%	82.6%

Less correct — More correct

Precision recall



Model evaluation measure

Model evaluation measure

Measures	Holdout score	Cross validation score
Accuracy	0.826	0.821
Area under ROC	0.763	0.759
Precision	0.706	0.697
Recall	0.364	0.336
F1	0.481	0.453
Average precision	0.541	0.540
Log loss	0.437	0.441

edunet foundation

CONCLUSION

The credit card defaulter prediction model achieved a high accuracy rate, successfully distinguishing between defaulters and non-defaulters. Using Snap Random Forest Classifier and IBM Watson Studio Auto AI, the model effectively handled imbalanced data and provided timely predictions. The results demonstrated reliable performance, supporting risk management and decision-making processes for financial institutions.

FUTURE SCOPE

- **Enhanced Data Sources:**
  •Incorporate additional data sources such as social media activity, transaction patterns, and economic indicators to improve model accuracy.

- **Real-Time Monitoring:**
  - Implement real-time monitoring and prediction systems to detect potential defaulters as early as possible, allowing for proactive risk management.

- **Personalized Customer Insights:**
  •Utilize the model to offer personalized financial advice and products to customers based on their risk profiles, enhancing customer experience and loyalty.

- **Integration with Financial Systems:**
  •Integrate the model with financial institutions' existing systems for automated decision-making in credit approvals, loan restructuring, and debt collection strategies.

REFERENCES

1. **Chen, T., & Guestrin, C. (2016).** *XGBoost: A scalable tree boosting system.* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.
2. **Brownlee, J. (2016).** *An Introduction to Gradient Boosting Decision Trees.* Machine Learning Mastery.
3. **UCI Machine Learning Repository. (n.d.).** *Default of Credit Card Clients Dataset.*
4. **IBM. (2023).** *AutoAI in Watson Studio: A no-code tool for automatic machine learning.* IBM Cloud Documentation.
5. **Kumar, R. (2019).** *Machine Learning Applications in Credit Risk Analysis.* Springer International Publishing.
6. **Kelleher, J. D., Namee, B. M., & D'Arcy, A. (2015).** *Fundamentals of Machine Learning for Predictive Data Analytics.* MIT Press.

COURSE CERTIFICATE 1

In recognition of the commitment to achieve professional excellence

Getting Started with Enterprise-grade AI

IBM

Joshwa Thomas

Has successfully satisfied the requirements for:

Getting Started with Enterprise-grade AI

Issued on: 13 JUL 2024

Issued by IBM

Verify: https://www.credly.com/go/Xwr1Ve8A

IBM

edunet foundation

COURSE CERTIFICATE 2

In recognition of the commitment to achieve professional excellence

Joshwa Thomas

Has successfully satisfied the requirements for:

Getting Started with Enterprise Data Science

Issued on: 13 JUL 2024

Issued by IBM

Verify: https://www.credly.com/go/FI5FiO9S

IBM

edunet foundation

THANK YOU