Enhancing Customer Churn Prediction Through Integrated Machine Learning and AI-Driven Dashboards: The Development of the TriView AI Platform

BY Hui Yi Huang

Business Problem

- Understanding **customer behavior** is crucial to improving retention
- Modern dashboards for data visualization often focus on **linear predictions**

Main Purpose

To streamline and accelerate decision-making by providing a single upgraded dashboard that delivers clearer insights and faster responses for churn analysis.

TRIVIEW AI DASHBOARD

1. Data visualization
2. Machine learning (ML)
3. Natural language generation (NLG)

KEY FEATURES FOR CHURN

Simplified and highlight essential factors

Dataset Features

'E-commerce Customer Churn Analysis and Prediction' from Kaggle.

Variable	Discerption
CustomerID	Unique customer ID
Churn	Churn Flag
Tenure	Tenure of customer in organization
PreferredLoginDevice	Preferred login device of customer
CityTier	City tier
WarehouseToHome	Distance in between warehouse to home of customer
PreferredPaymentMode	Preferred payment method of customer
Gender	Gender of customer
HourSpendOnApp	Number of hours spend on mobile application or website
NumberOfDeviceRegistered	Total number of deceives is registered on particular customer
PreferedOrderCat	Preferred order category of customer in last month
SatisfactionScore	Satisfactory score of customer on service
MaritalStatus	Marital status of customer
NumberOfAddress	Total number of added added on particular customer
Complain	Any complaint has been raised in last month
OrderAmountHikeFromlastYear	Percentage increases in order from last year
CouponUsed	Total number of coupon has been used in last month
OrderCount	Total number of orders has been places in last month
DaySinceLastOrder	Day Since last order by customer
CashbackAmount	Average cashback in last month

- E-commerce focus
- Manageable complexity for light cleaning
- Small enough to prioritize development
- Real-world data, ensuring applicability
- Enough feature variety

Techniques

Tableau's Function

- Integration with Python – **TabPy**
- Extension of Generative AI – **Arria**
- Real-time dynamic prediction and report

Machine Learning

Classification Models:

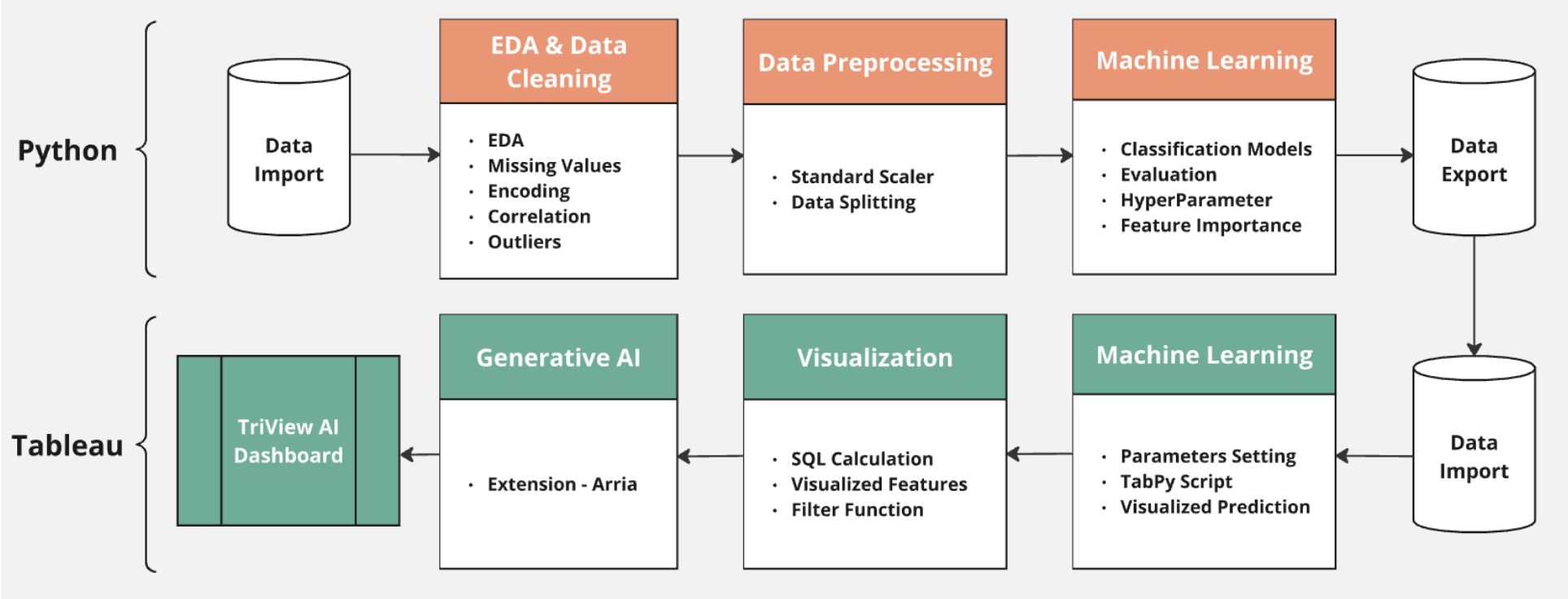
Logistic Regression, Support Vector Classifier (SVC), Multilayer Perceptron (MLP), Random Forest, Decision Tree, Gradient Boosting.

Limitation

TabPy's Trade-off

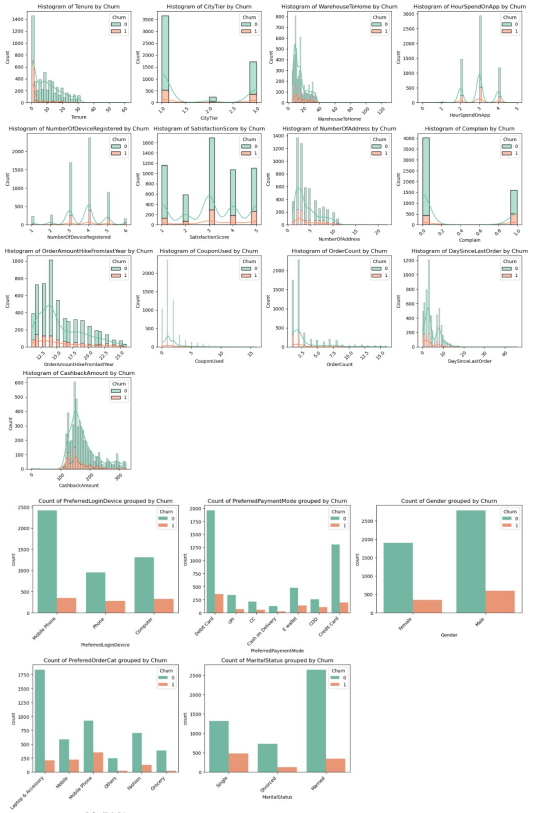
Balancing Efficiency and Complexity

Pipeline



Python – EDA (Exploratory Data Analysis)

- **Summary statistics** for key variables
- **Distributions** of numerical features and categorical features by histograms, segmenting the chart into churned and non-churned customer groups.
- Examined **correlations** between features



Key Insights

- **Tenure:** Customers with shorter tenure have higher churn rates.
- **Complain:** Customers who complaints are more likely to churn.
- **Marital Status:** Single customers show a higher likelihood of churning.
- **Satisfaction Score:** High satisfaction alone may not guarantee retention.
- **Day Since Last Order**, **Number Of Device Registered**, and **Cashback Amount** have minor correlation with churn.

Python – Preparing for Machine learning

Data Cleaning & Preparation

- Imputed **missing values** (median/mode).
- One-Hot **Encoding** for categorical variables.
- **Outlier** treatment with Interquartile Range (IQR) method.
- **Standardized** features and **split** data.

Modeling & Evaluation

- Implemented selected **classification models**
- **Evaluated** models using accuracy, F1-score, and recall score.

	Model	DataFrame	Accuracy	F1 Score	Recall Score
0	Logistic Regression	Standardized	0.891285	0.623053	0.512821
1	SVC	Standardized	0.911051	0.687697	0.558974
2	MLP	Standardized	0.964061	0.893617	0.861538
3	Random Forest	Standardized	0.955076	0.860335	0.789744
4	Decision Tree	Standardized	0.957772	0.881612	0.897436
5	Gradient Boosting	Standardized	0.907457	0.696165	0.605128

Python – Best models, Feature Importance

- Used Cross-Validation to fine-tune **hyperparameters** and achieved optimal models.
- Extracted Feature Importance based on the models, MLP and Random Forest (Recall score: 0.85).

Key Insights

- **Number of Addresses, Warehouse to Home,** and **City Tier,** which did not appear as significant in the earlier analyses.

	Top Feature
0	Tenure
1	Complain
2	DaySinceLastOrder
3	SatisfactionScore
4	NumberOfAddress
5	NumberOfDeviceRegistered
6	WarehouseToHome
7	CashbackAmount
8	MaritalStatus_Single
9	CityTier

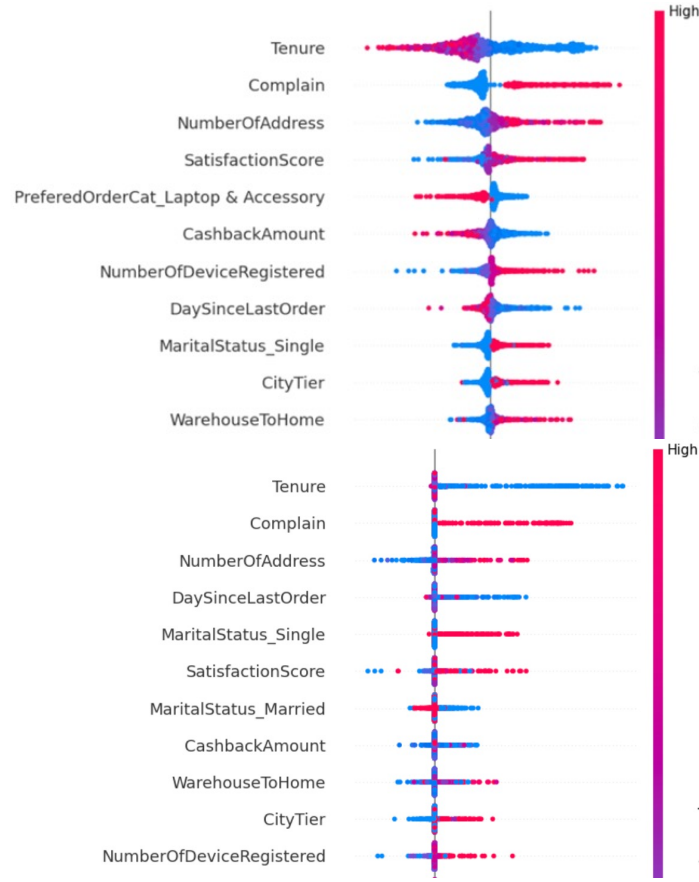


Tableau – TabPy for Machine Learning

Machine Learning

- **Exported** the dataset into a list with all features in same data types for TabPy.
- Defined the top 10 features as **parameters**, applied **StandardScaler**, and integrated the **Random Forest Classifier** with the best parameters.
- Created a **dynamic visualization** that displays real-time churn predictions.

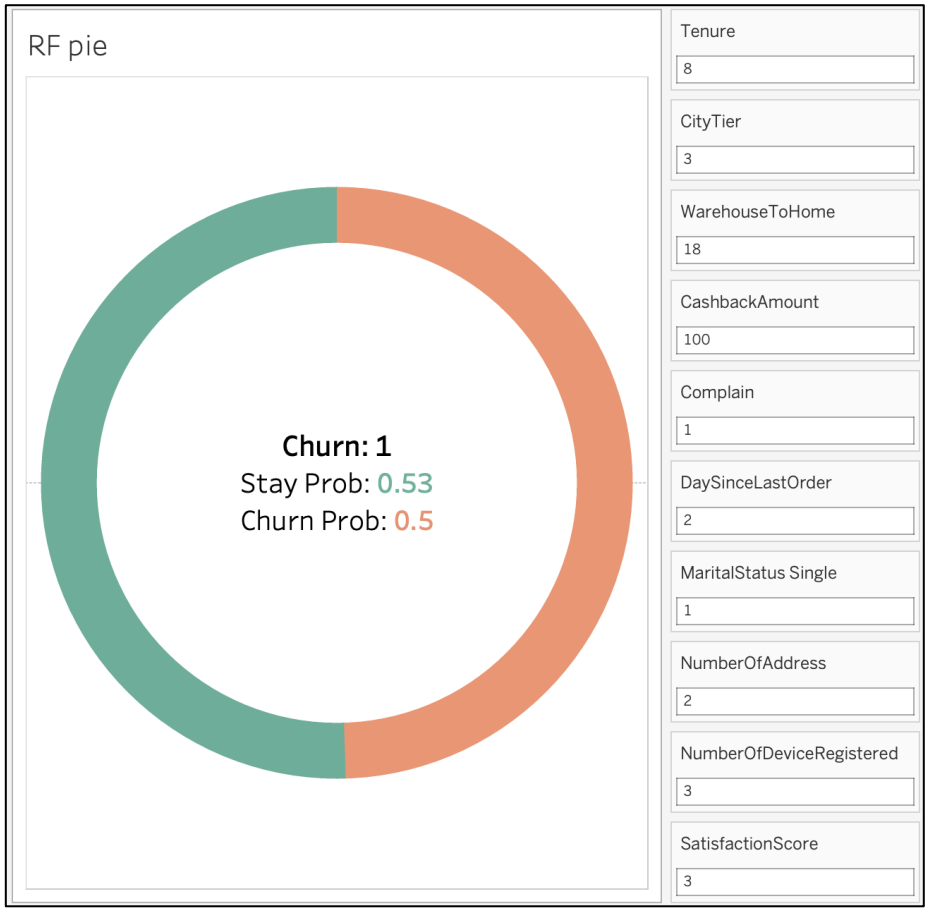


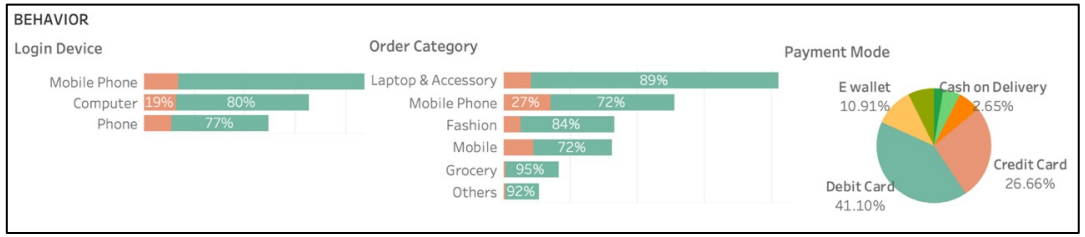
Tableau – Visualizations and Generative AI

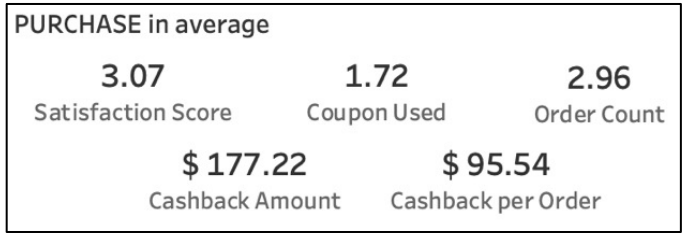
Visualizations

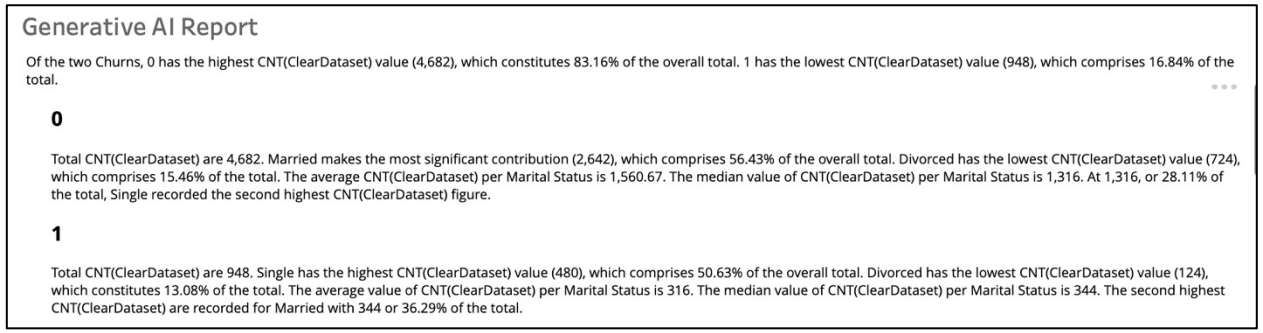
- Created key metrics using SQL calculated fields.
- Divided into demographic section, behavior section, activity section, and purchase section.
- Created dynamic visualizations with interactive filters of churn value.

Generative AI

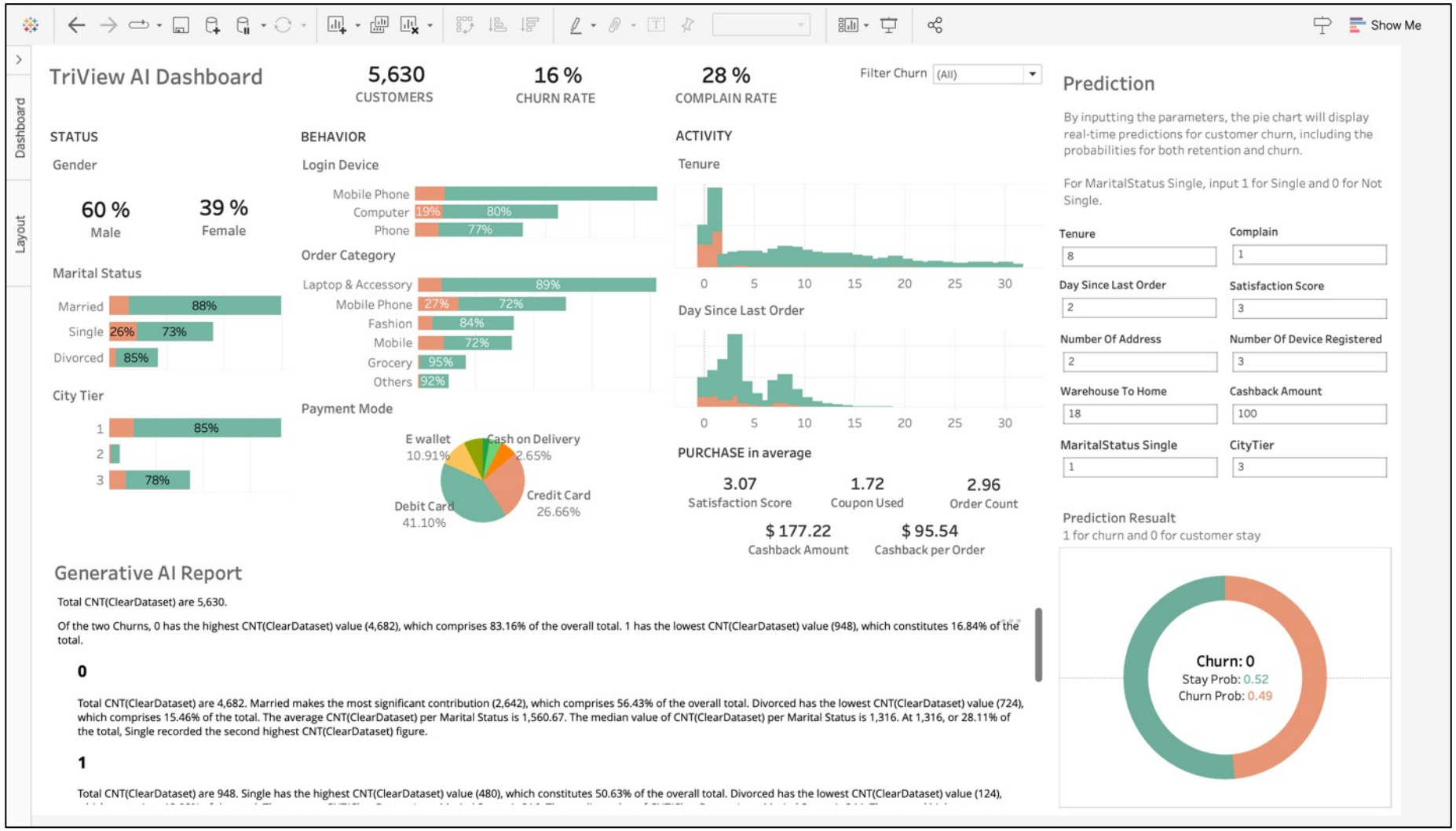
- Integrated **Arria**, the extension of generative AI, for automated reports based on user-selected data.







TriView AI Dashboard

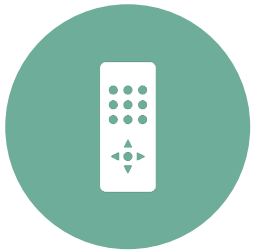


Objective Achievements

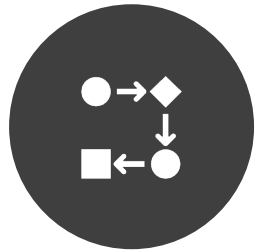
- Uncovered the crucial hidden features that significantly influence churn rates.
- The integration of machine learning and natural language generation within a Dashboard.

Benefits of TriView AI Dashboard for Business

REAL-TIME PREDICTION



IDENTIFICATION OF KEY FEATURES



STREAMLINED DECISION-MAKING

Q&A

BY Hui Yi Huang