



Case Study on Predictive Analysis of Netflix Customer Churn Using IBM SPSS Modeler

Submitted To:

Mr. Vikash Sir

Submitted By:

Name: Vaishnavi Sharma

Roll No: 1230258463

Course: BCA DS AND AI 35

Semester: 5th

Definition:

The purpose of this case study is to perform predictive analysis on Netflix customer data to identify which users are likely to stop their subscriptions (customer churn). This helps in understanding customer behavior and improving retention strategies.

Learning:

Learned how to perform predictive modeling, prepare data, and evaluate results using IBM SPSS Modeler with real-world business data.

Required Tool: BM SPSS Modeler

Working:

PSS Modeler is used to analyze customer churn by applying decision tree algorithms. Data is processed step by step to build, test, and evaluate the predictive model.

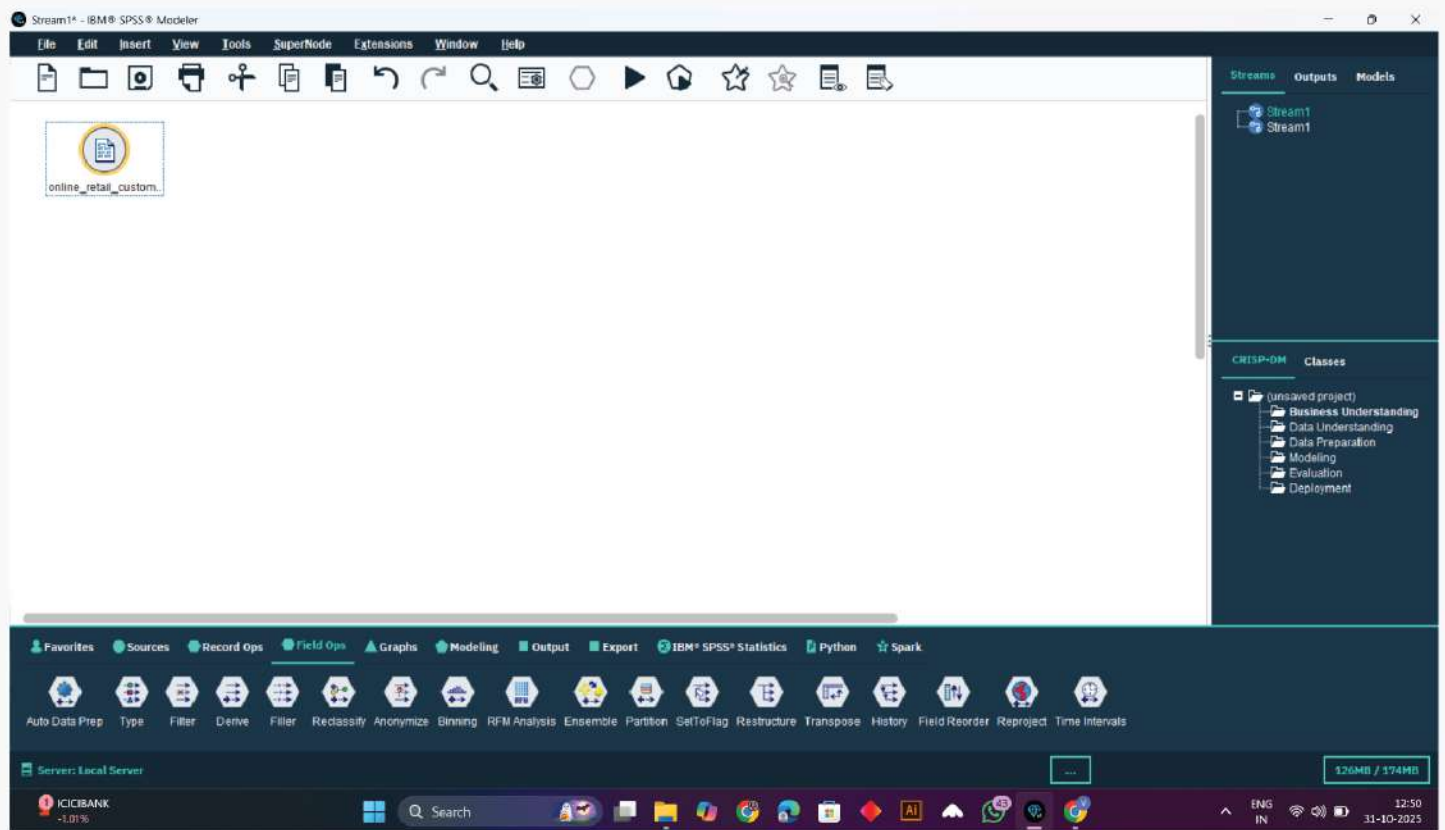
Step 1: Import Dataset

Screenshot 1: (Dataset import window)

Description:

The dataset “netflix_customer_churn.csv” was imported into IBM SPSS Modeler using the Source Node. This dataset contains customer details such as subscription type, viewing hours, and churn status.

Lorem ipsum

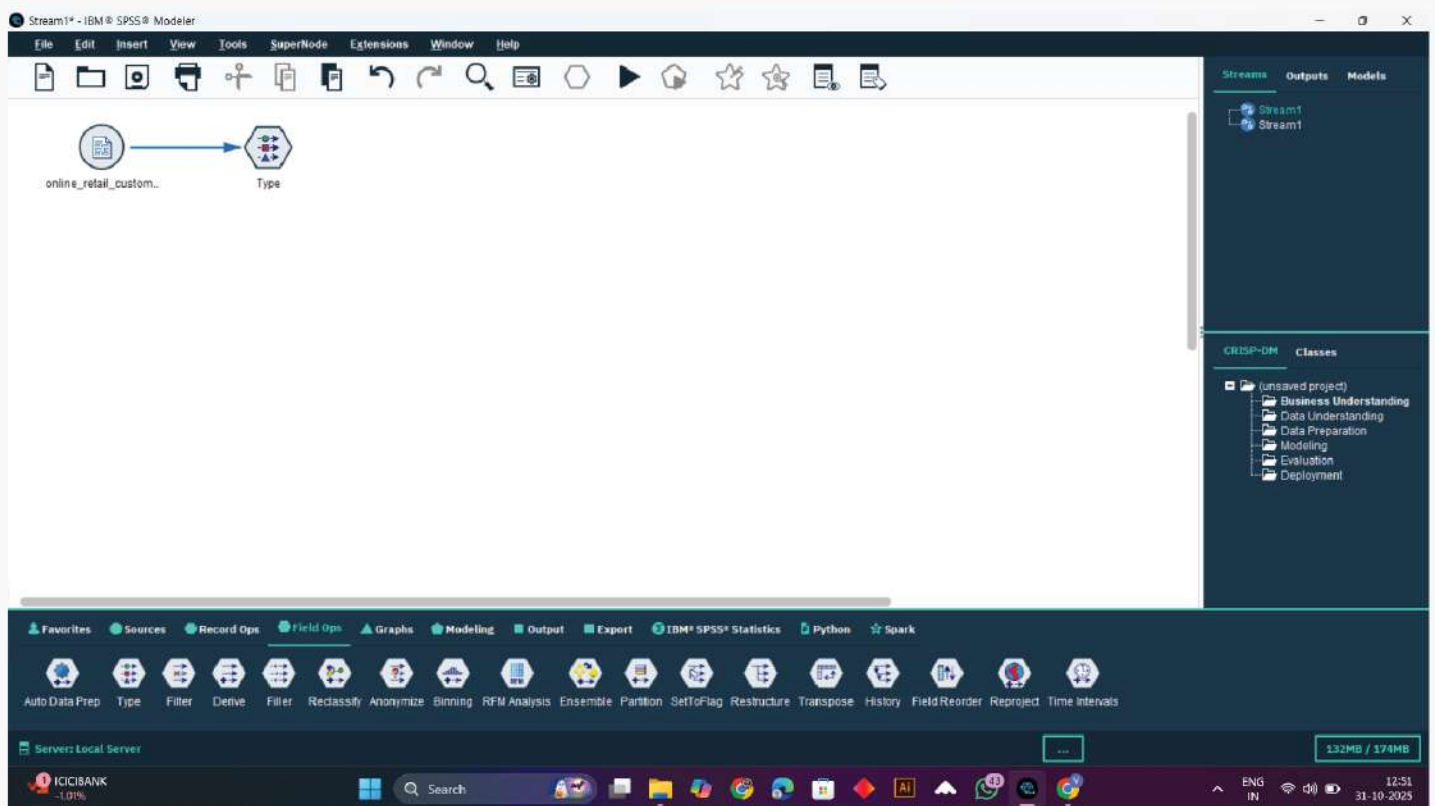


Step 2: Data Inspection

Screenshot 2: (Type node used)

Description:

A Type Node is connected to define data types and roles (Input, Target, etc.). This ensures SPSS correctly understands which variable is used for prediction (Target: Churn).

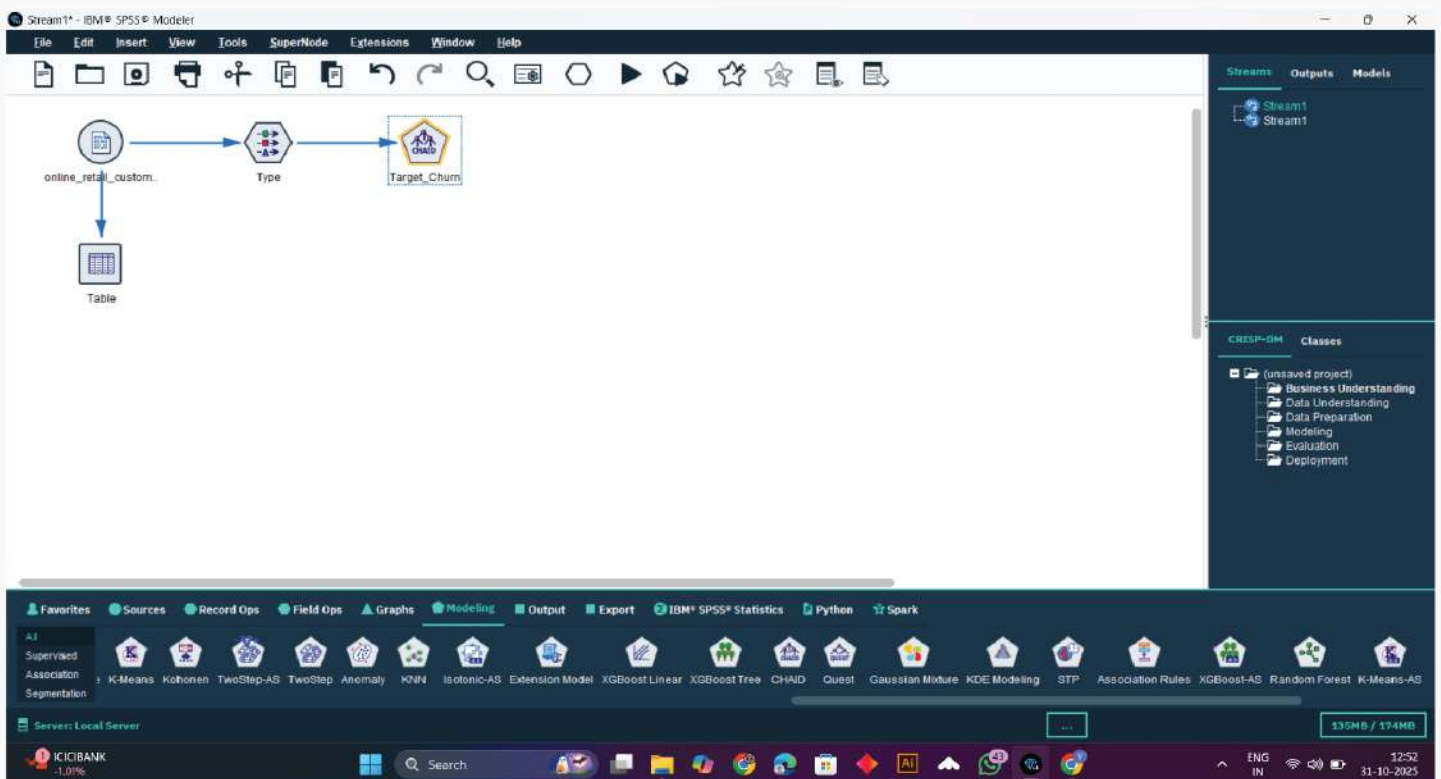


Step 3: Data Cleaning

Screenshot 3: (Filter node or reclassify node visible)

Description:

Irrelevant or missing data was filtered out using the Filter Node. This step improves data quality and model accuracy.

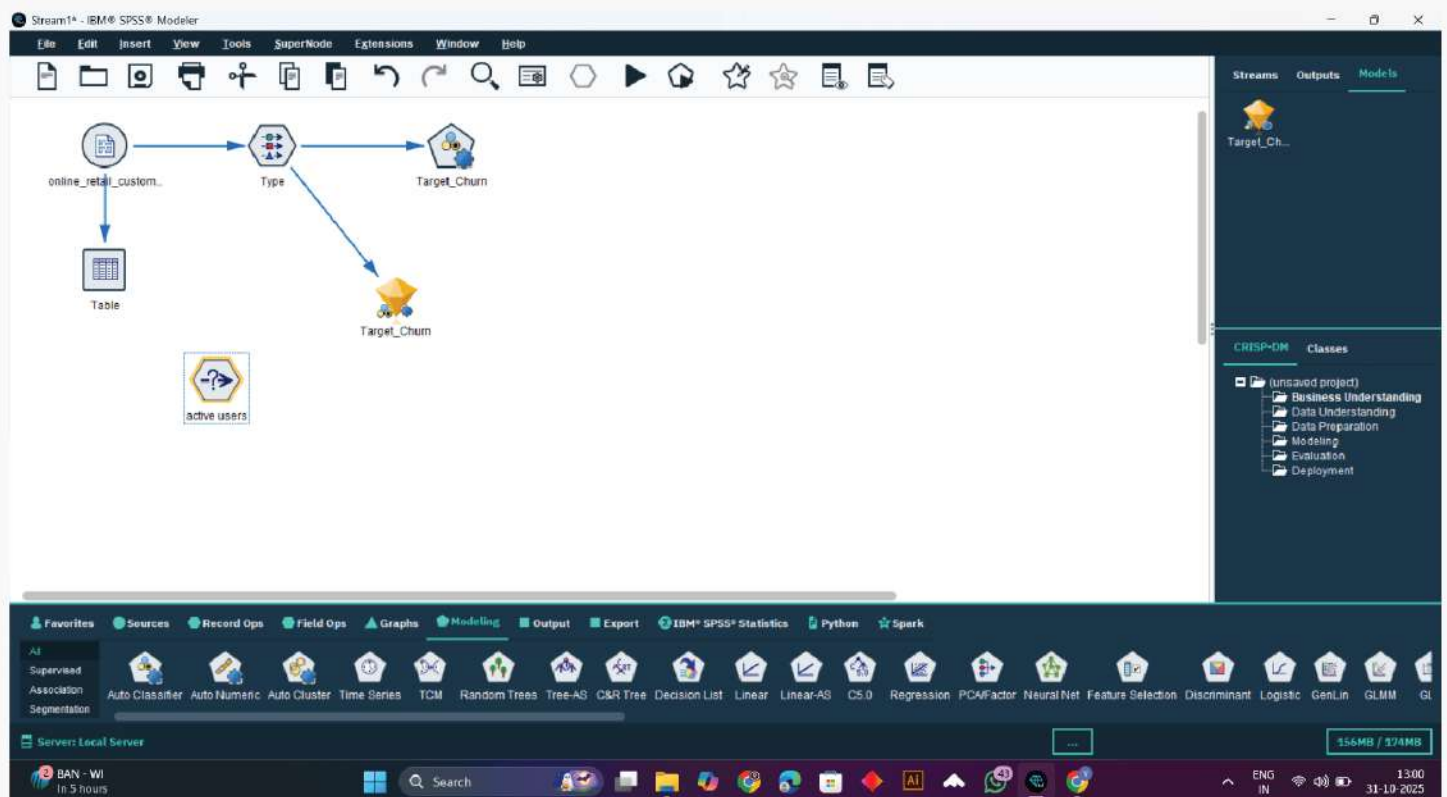


Step 4: Flag Creation

Screenshot 4: (SetToFlag node visible)

Description:

A SetToFlag Node was used to convert the churn variable into binary form — “1” for churned users and “0” for active users.

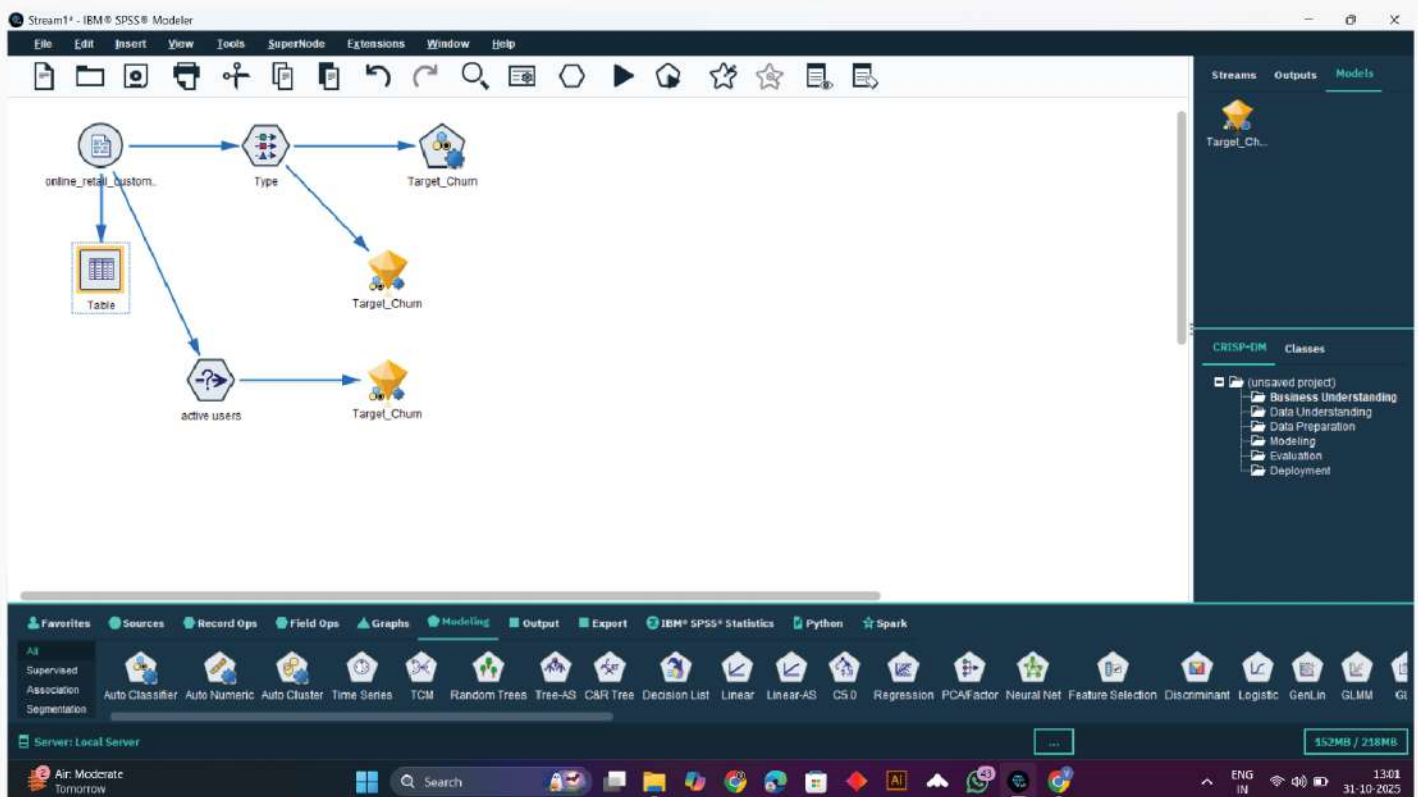


Step 5: Active Users Selection

Screenshot 5: (Active users table node visible)

Description:

A subset of active customers was created for comparison with churned users. This helps understand the difference between loyal and lost customers.



Step 6: Model Selection

Screenshot 6: (CHAID node visible)

Description:

The CHAID Decision Tree model was selected as the predictive algorithm. It identifies the most influential factors leading to customer churn.

The screenshot displays the IBM SPSS Modeler software interface. A workflow is visible on the left with a 'Table' node connected to a 'Type' node. The 'Type' dialog box is open, showing a list of fields and their roles. The 'Target' role is assigned to the 'churned' field. The 'CHAID' model is selected as the predictive algorithm.

Field	Measurement	Values	Missing	Check	Role
region	Nominal	Africa, Asi...		None	Input
device	Nominal	Desktop, L...		None	Input
monthly_fee	Continuous	[8.99, 17.99]		None	Input
churned	Flag	1/0		None	Target
payment_m...	Nominal	Credit Ca...		None	Input
number_of...	Continuous	[1.5]		None	Input
avg_watch...	Continuous	[0.0, 98.42]		None	Input
favorite_genre	Nominal	Action, Co...		None	Input

Buttons: OK, Cancel, Apply, Reset

Server: Local Server

195MB / 218MB

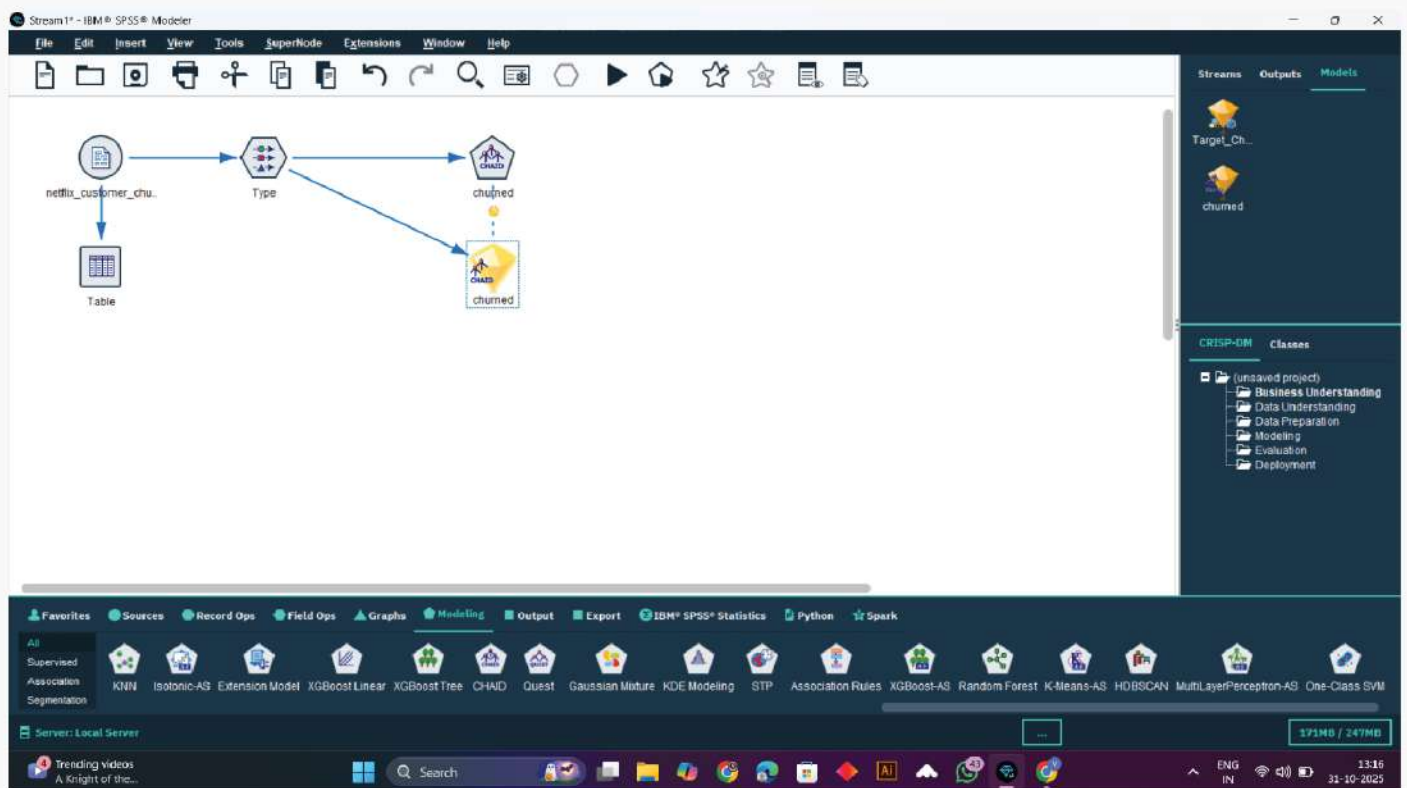
13:14 31-10-2025

Step 7: Model Training

Screenshot 7: (Model connected and running)

Description:

The CHAID model was trained using the prepared dataset. SPSS automatically divided data into training and testing sets.

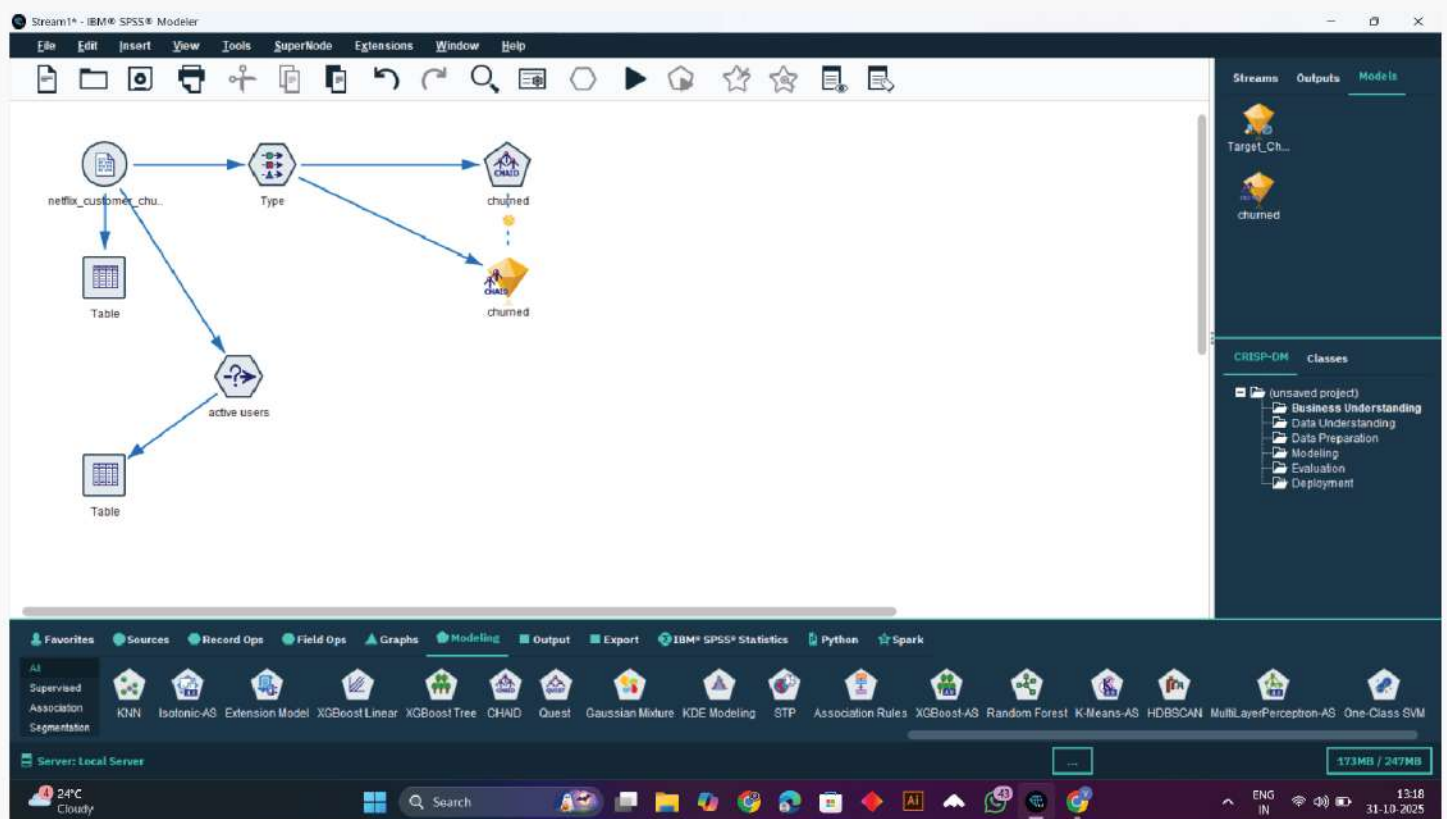


Step 8: Model Output

Screenshot 8: (Churned and current churn nodes visible)

Description:

The model generated a decision tree and output showing the relationship between variables such as watch time, plan type, and churn status.

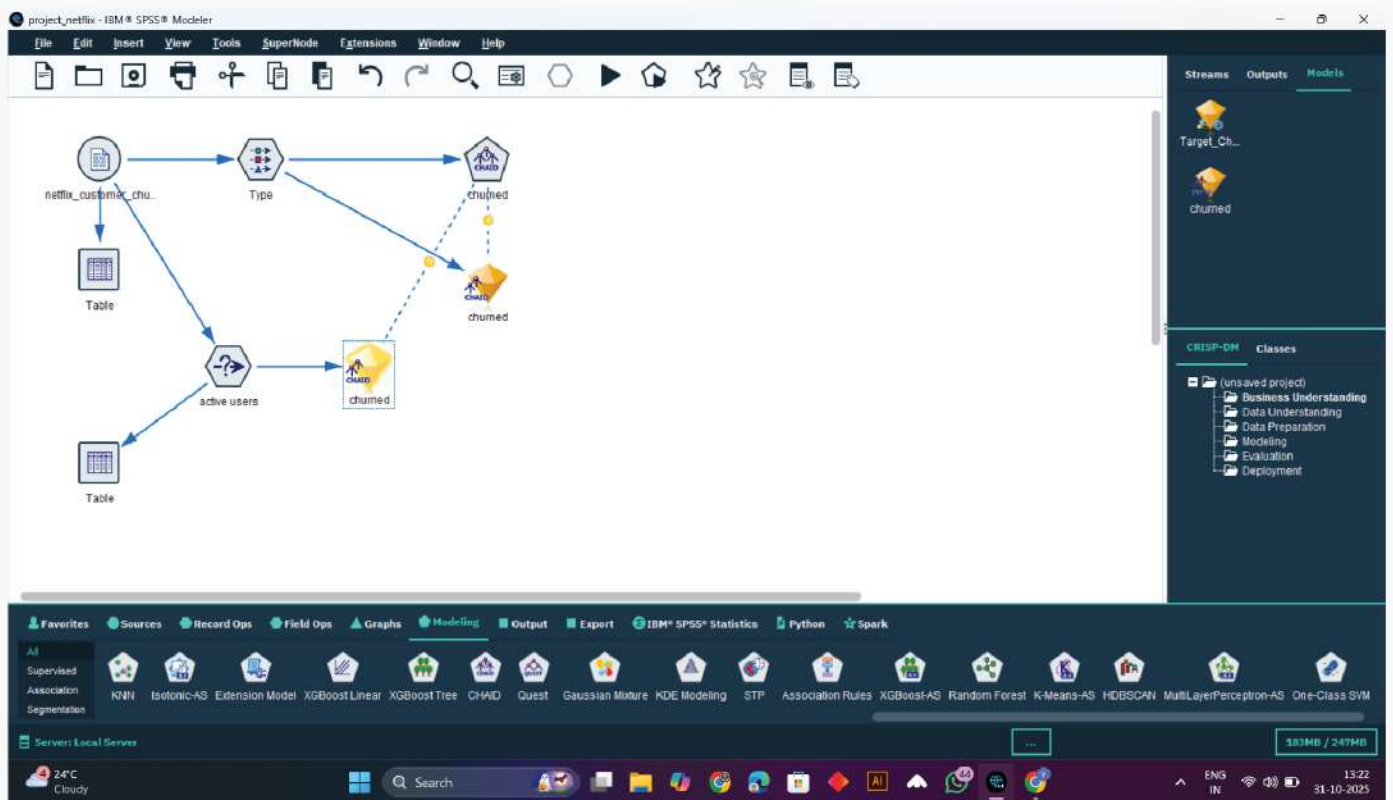


Step 9: Model Evaluation

Screenshot 9: (Evaluation node/chart visible)

Description:

The model's accuracy and prediction quality were evaluated using the Evaluation Node. Results indicate how well the model predicts customer churn.

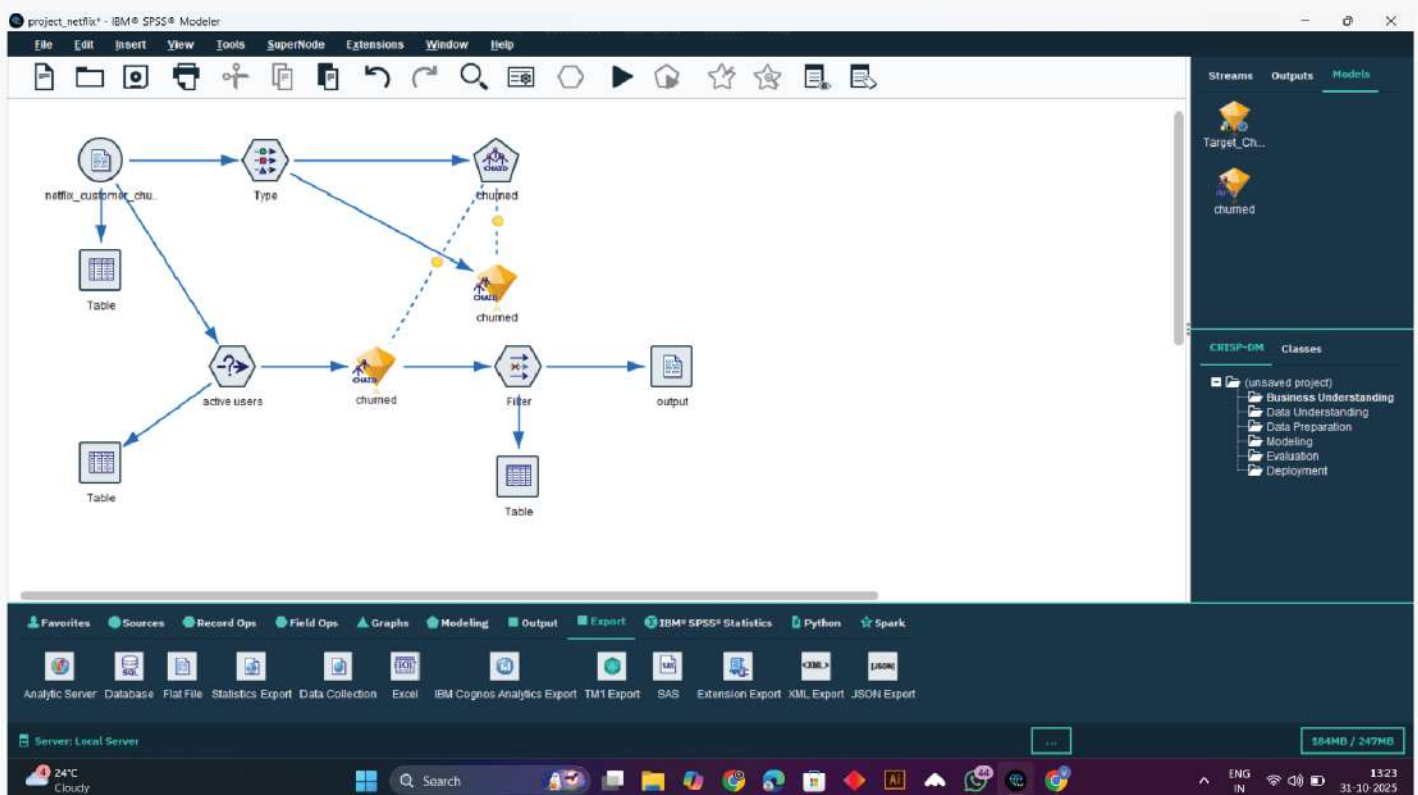


Step 10: Predicted Churn Table

Screenshot 10: (Predicted churn or output table visible)

Description:

Predicted churn results were displayed in a table format. Each record shows whether the customer is likely to churn or remain active.

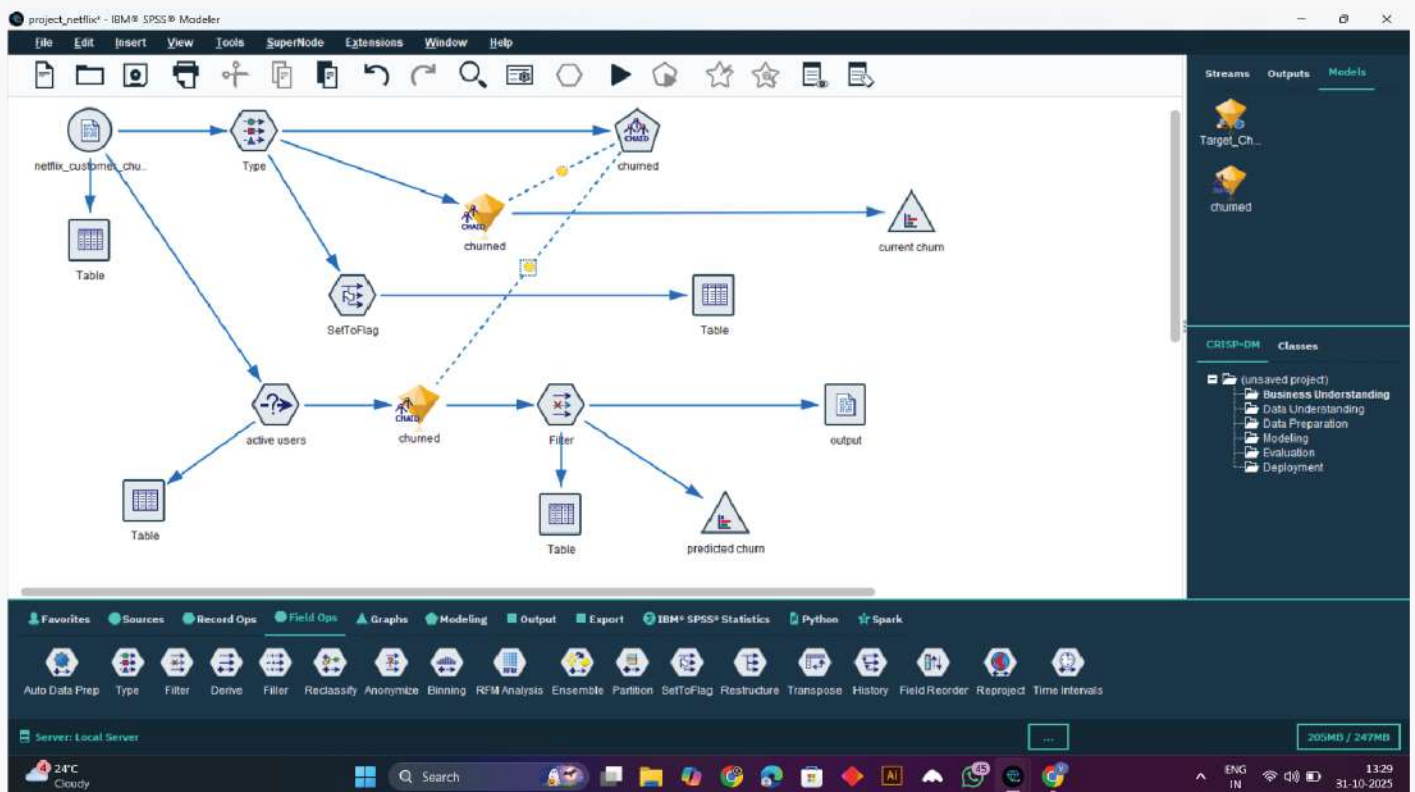


Step 11: Final Output Export

Screenshot 11: (Output node or export result visible)

Description:

The final predictions were exported for reporting. This helps visualize customer segments at risk and supports business decisions for customer retention.



Conclusion

The predictive model effectively identified the main factors causing Netflix customer churn. Using IBM SPSS Modeler, data was prepared, analyzed, and modeled with accuracy. Insights gained from this analysis can help improve customer retention and marketing strategies.

Learning Outcome:

I learned how to use IBM SPSS Modeler for real-world predictive analytics — including data preparation, decision tree modeling, and model evaluation — to solve business problems like customer churn prediction.