

# Babu Banarasi Das University



## CASE STUDY ON Predicting Student Churn Using IBM SPSS Modeler

SUBMITTED TO: MR. AYUSHMAN BHADURIA

SUBMITTED BY: Chandra Bhan Maurya

# **Agenda:**

To analyze student academic and behavioral data in order to develop a predictive model that identifies students who are likely to drop out (churn) from the institution.

# **Outcomes:**

A working predictive model that helps the institution detect at-risk students early, enabling timely academic or counseling interventions to reduce dropout rates.

# **Required Tools:**

IBM SPSS Modeler for data processing and model building, MS Excel for dataset handling.

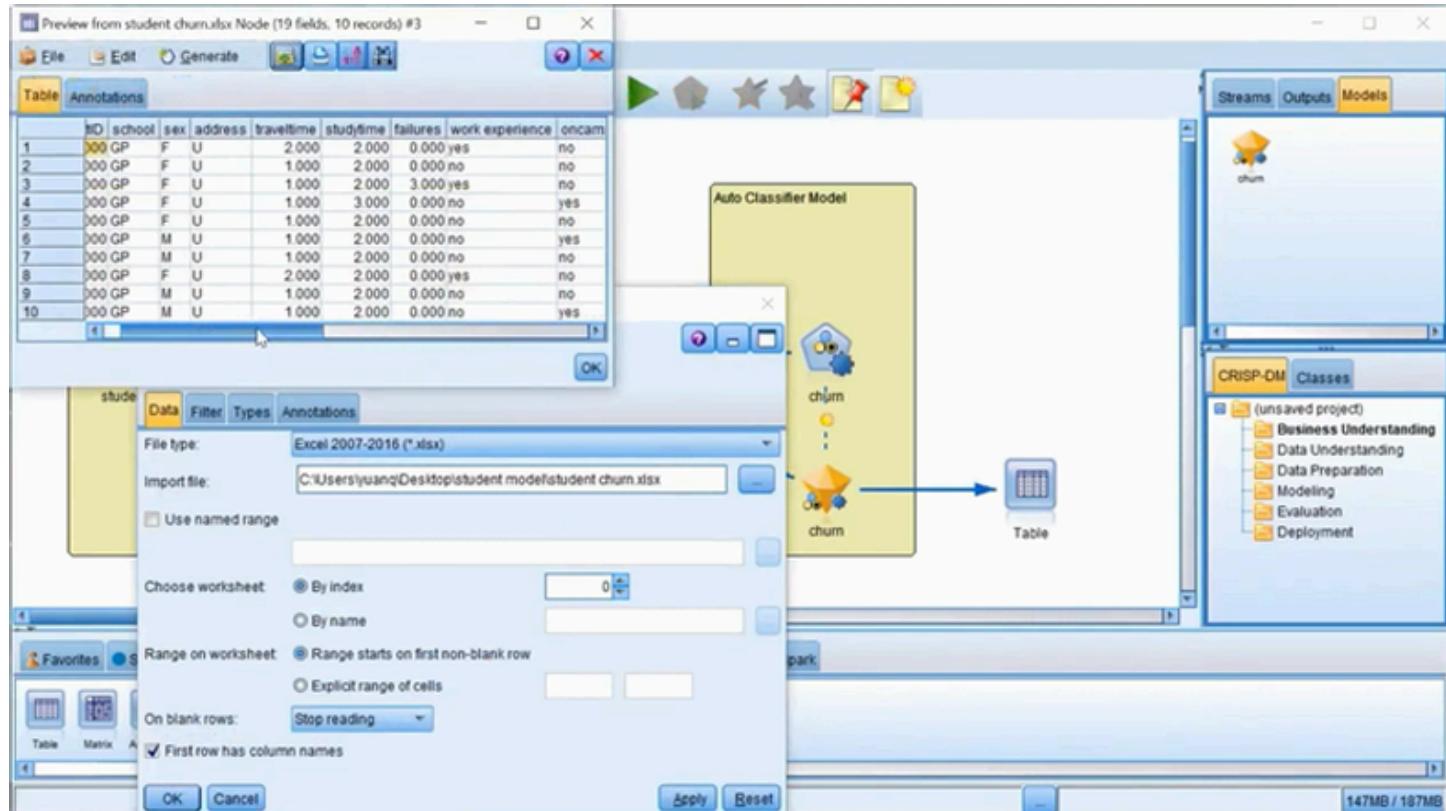
# **Working:**

Collect student data → Import and preprocess it in IBM SPSS Modeler → Build and evaluate predictive models → Select the most accurate model → Interpret results and recommend intervention strategies.

# Steps to Perform the Project

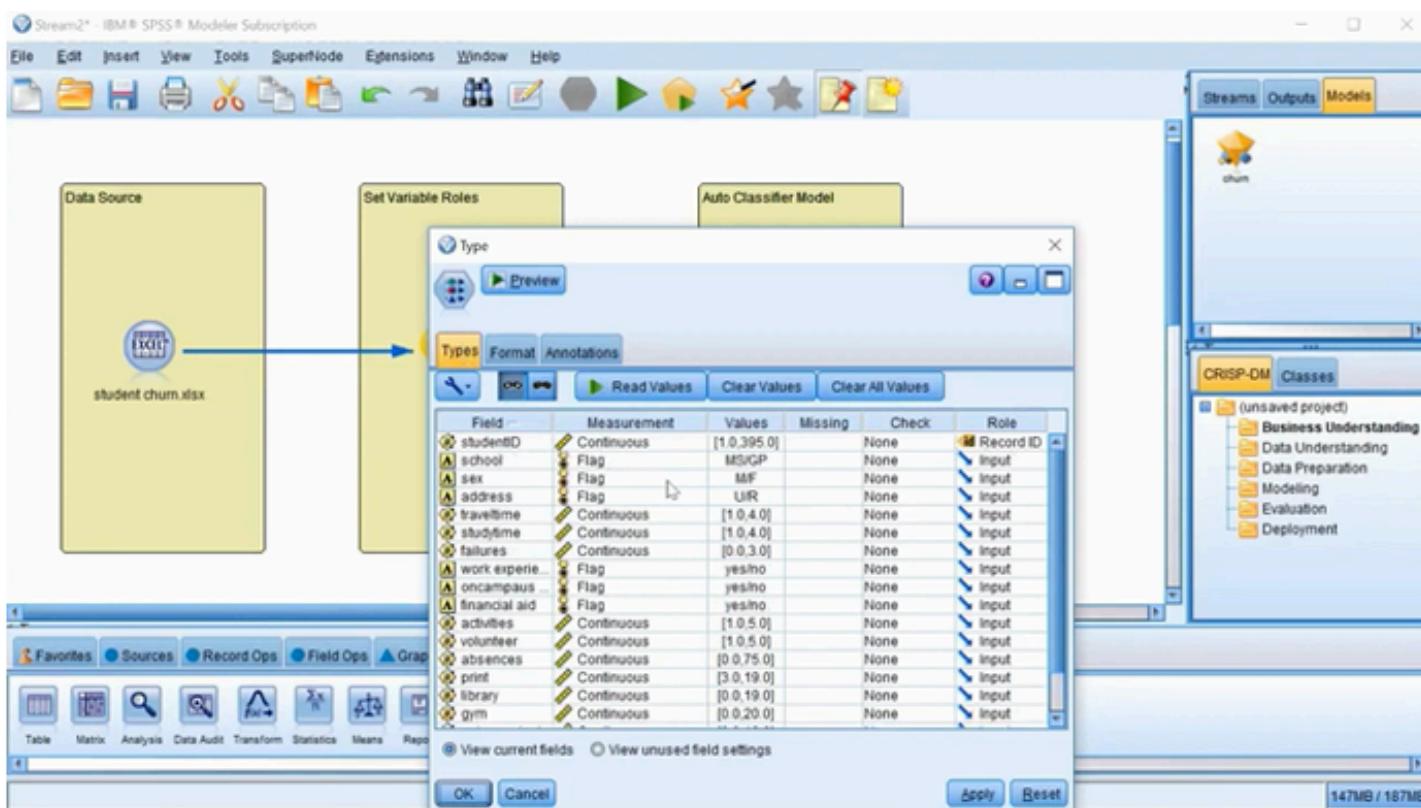
## Step 1: Import the Dataset:

Open IBM SPSS Modeler and bring in the student dataset containing academic, attendance, and demographic records



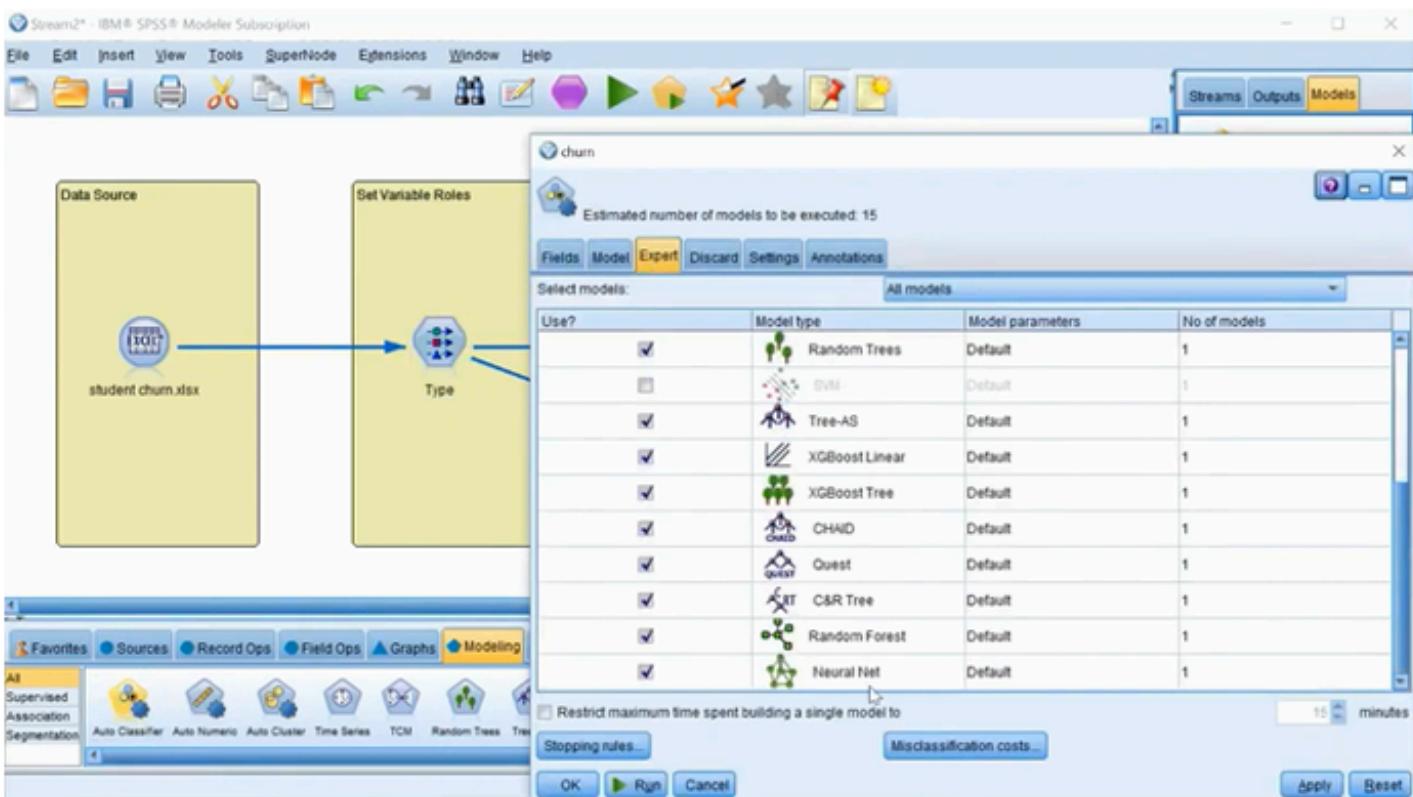
## Step 2:

Set Variable Roles: Assign Student Status as the Target variable and mark other related fields such as marks, attendance, activities, etc., as Input variables.



## Step 3: Select the Model:

Drag and place the Auto Classifier node to automatically test and compare multiple classification algorithms.



The screenshot shows the IBM SPSS Modeler interface. On the left, there is a 'Data Source' node labeled 'student\_churn.xlsx'. An arrow points from this node to a 'Set Variable Roles' node. From the 'Set Variable Roles' node, another arrow points to an 'Auto Classifier Model' node. A dialog box titled 'churn' is open, showing three models: Bayesian Network 1, XGBoost Tree 1, and Random Trees 1. The 'Model' tab is selected, displaying details like build time, max profit, and overall accuracy. The 'XGBoost Tree 1' model has the highest accuracy at 90.886. The interface also includes a toolbar at the top and a navigation bar at the bottom.

## Step 4: Run the Model:

Execute the Auto Classifier to generate the best performing model based on accuracy and classification performance.

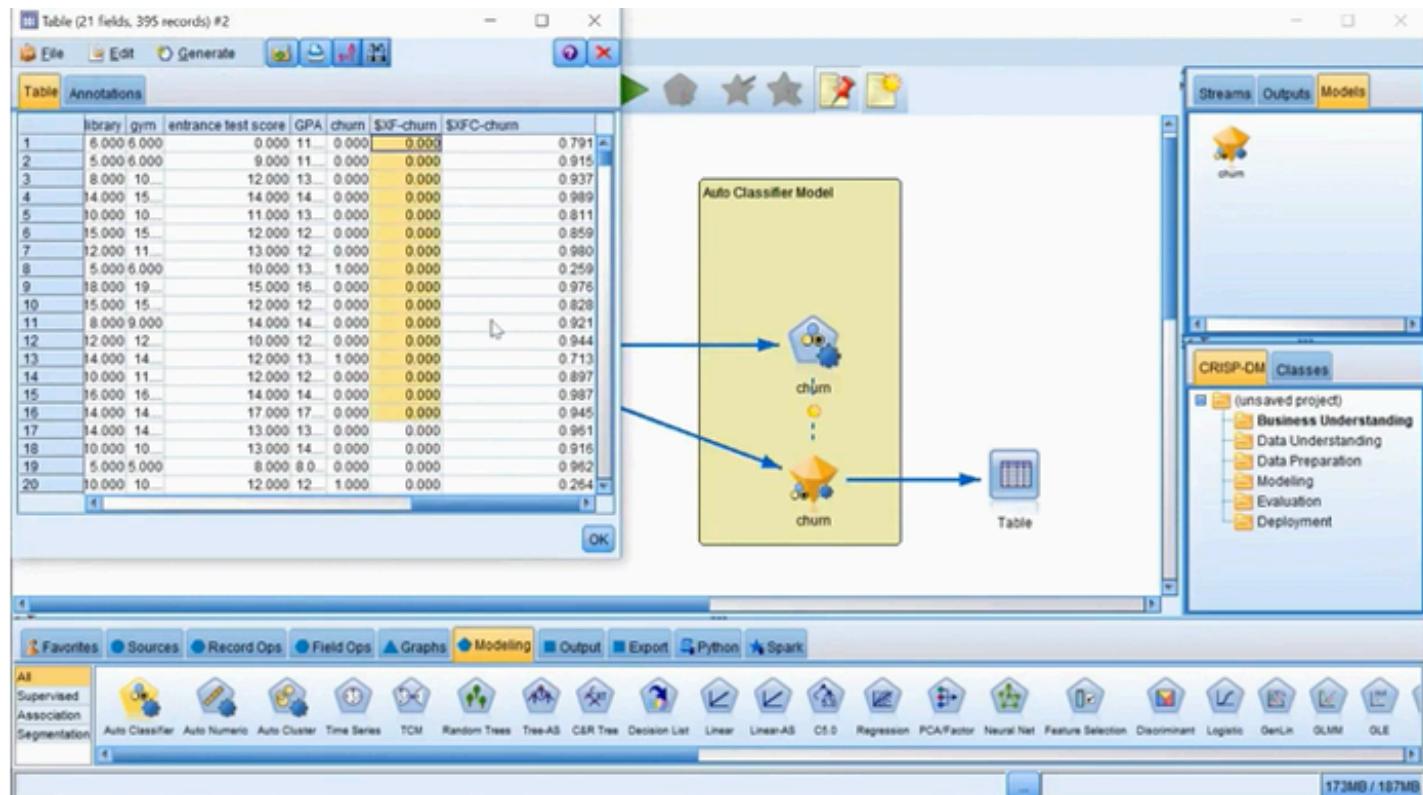
The screenshot shows the 'Bayesian Network' dialog box. It displays a directed acyclic graph (DAG) representing causal relationships between variables: 'churn' is the target variable (blue circle), influenced by 'volunteer', 'sex', 'studytime', 'failures', 'print', 'work-experience', 'address', 'school', 'entrance-test score', 'GPA', 'gym', 'library', 'absences', 'oncampus job', 'traveltime', and 'activities'. The 'Conditional Probabilities of school' table shows the probability of being in class 1 ('MS') or class 0 ('GP') given different values of 'churn'. The table is as follows:

Parents	Probability	Probability
churn	MS	GP
1	0.11	0.89
0	0.12	0.88

The interface also includes a 'CRISP-DM' panel on the right showing the project structure, and a toolbar at the bottom.

# Step 5: Evaluate and Interpret Results:

Review charts, accuracy tables, and predictor importance to understand which factors influence student churn.



## Conclusion:

The project successfully demonstrated how student churn can be predicted using IBM SPSS Modeler. By defining the target variable and using the Auto Classifier model, we identified the key academic and behavioral factors that influence whether a student continues or drops out. The predictive results help institutions take early action, such as providing counselling, academic support, or attendance monitoring, to reduce dropout rates and improve overall student retention.