Babu Banarasi Das University

BBD UNIVERSITY

Case Study On

Bank Churn Prediction using SPSS Modeler

SUBMITTED TO: **SUBMITTED BY:**

 Mr. Vikas Sir   **Name: Lavanya & Team**

Bank Churn Prediction — SPSS Modeler Project

Purpose: Build, evaluate and deploy churn model using the provided bank_churn.csv dataset. This file is a step-by-step project guide to follow inside IBM SPSS Modeler to produce the required deliverables.

---

1. Project overview

- **Goal:** Predict whether a customer will churn (Exited = 1) using available features (Age, Balance, CreditScore, Geography, Gender, etc.).

- **Train/test split:** 70% train / 30% test (random)

---

2. Files used

- bank_churn.csv — original dataset (10,000 rows)

- bank_churn_train.csv — 70% random split (for training)

- bank_churn_test.csv — 30% random split (for testing / scoring)

---

3. Stream A — Model building (training & validation)

Var File (bank_churn_train.csv)

  → Type Node

 →CHAID Node

  → Generate Model (Model Nugget)

Stream B — Deployment

Var File (bank_churn_test.csv)

  → Generated Node

 →Filter Node

  → Model Nugget (drag gold cube from Stream A)
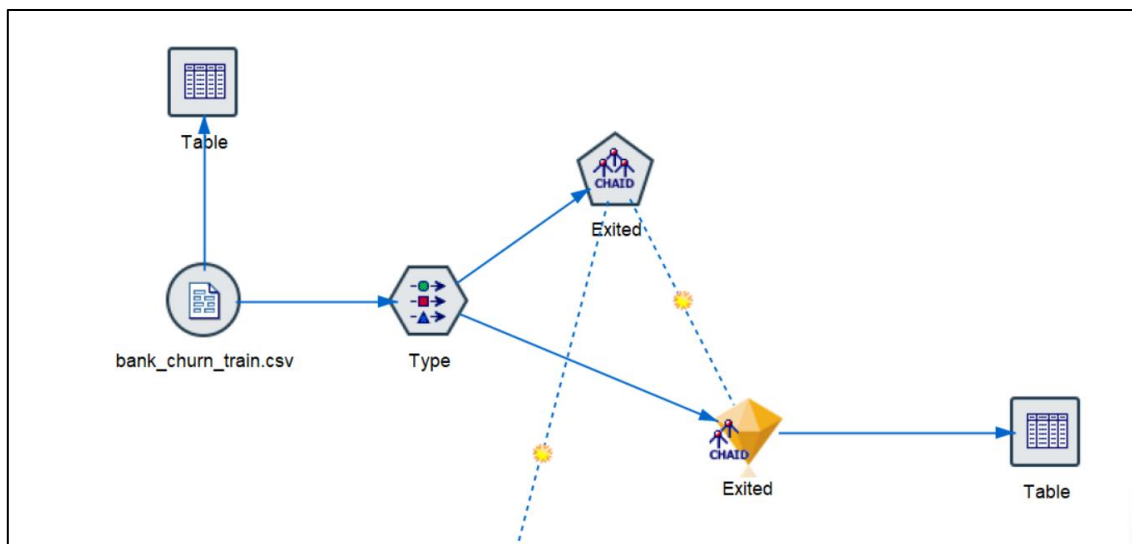
  → Table Node

---

4. stream A

1. **Create new stream** in SPSS Modeler; save it as BankChurn_Project.str.

2. **Var. File node**

   o Drag Var. File node and open bank_churn_train.csv.

   o In file options, ensure the delimiter is comma and header row is used.

3. **Type node** (important)

   o Add a Type node immediately after Var. File.

   o Set field roles and measurement levels:

      ▪ Exited → Role = Target, Type = Nominal (or Binary)

      ▪ Geography → Type = Nominal

      ▪ Gender → Type = Nominal

      ▪ HasCrCard, IsActiveMember, NumOfProducts → Type = Ordinal or Numeric, Role = Input

      ▪ RowNumber, CustomerId, Surname → Role = None (these are identifiers)

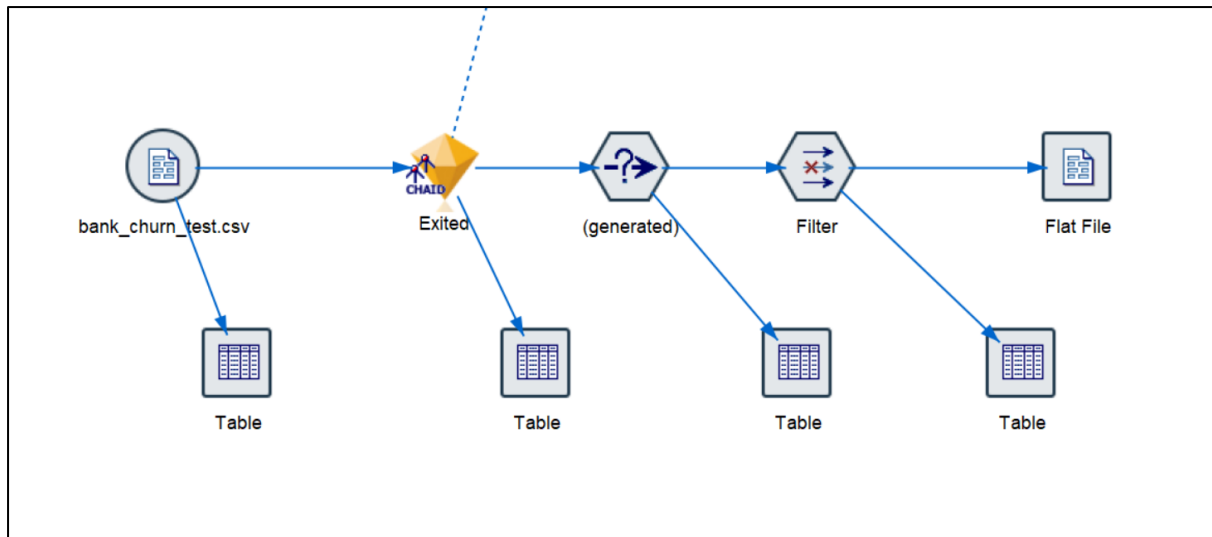   o Check for missing values and correct data types.

4. **Generate model nugget**

   o Run the CHAID model which will create the gold cube nugget which contains the trained model and connect it with a table node and run to see the output or result.

---

5. stream B

1. Create a new Var. File node and load bank_churn_test.csv.

2. **Model Nugget**: drag the gold nugget (created in Stream A) to Stream B.

3. **Connect**: Var. File (test) → Model Nugget → Select node→ Table node.

4. **Run** the Table node. Inspect the output columns.

8. Troubleshooting common issues

- **"Unexpected end-of-file" / "Incorrect number of fields"** on CSV import:

    o Open CSV in Excel, remove trailing blank rows, ensure every row has the same number of columns separated by commas, then save as CSV.

- **Field type mismatches**:

    o Use Type node to force correct data types (Nominal, Continuous, Date).

- **Model overfitting**:

    o Use Partition node and cross-validation or use regularization in logistic regression (if supported) or simplify predictors.