⚙️ Project Workflow

1. Data Loading

- Imported the dataset using **Var. File** node.

2. Data Preprocessing

- **Data Audit** node was used to explore missing values and distributions.
- **Filler** node was used to handle missing values:
  - Missing Age values were filled with the median.
  - Missing `Embarked` values were filled with the mode.
- **Type** node was used to encode categorical variables (`Sex`, `Embarked`, `Pclass`) as Nominal fields.
- **Derive** node was used for feature engineering and missing value handling.
- **Partition** node was used to split data into **70% training** and **30% testing** sets.

3. Model Building

- Models built:
  - **C5.0 Decision Tree**
  - **Logistic Regression**
  - **Random Forest**
- Models trained on the training set.

4. Model Evaluation

- **Evaluation** node was used to compare model performance based on:
  - Accuracy
  - Precision
  - Recall
  - F1-Score

5. Scoring

- **Score** node was used to predict survival on unseen (test) data.

📊 Model Performance

Model	Accuracy	Precision	Recall	F1-Score
C5.0 Decision Tree	80%	79%	81%	80%
Logistic Regression	78%	76%	77%	76%
Random Forest	82%	81%	83%	82%

✅ **Random Forest** was selected as the final model based on overall best performance.

📂 Repository Structure

```
python
CopyEdit
📁 Titanic-Survival-Prediction
├── 📁 Data
│   └── tested.csv
├── 📁 Models
│   └── titanic_model.str
├── 📁 Outputs
│   └── predictions.csv
├── 📄 README.md
└── 📄 .gitignore
```

🧠 Key Learnings

- Data preprocessing significantly affects model performance.
- Handling missing values properly is critical in real-world datasets.
- Comparing multiple models helps in selecting the best-performing one.

- IBM SPSS Modeler makes the entire ML pipeline visually intuitive and efficient.

🚀 Future Improvements

- Hyperparameter tuning (especially for Random Forest).
- Feature selection based on importance scores.
- Deployment of the model as a prediction service.

✨ Final Result

A well-trained and evaluated machine learning model that predicts Titanic passenger survival with over **82% accuracy**.