Customer Churn Prediction Challenge

Overview:
Customer churn is a significant concern for many industries, particularly in the telecommunications sector, where retaining customers is often more cost-effective than acquiring new ones. This challenge focuses on using the **Telco Customer Churn Dataset** to build a KNIME workflow that predicts customer churn. Participants will utilize KNIME's suite of nodes to perform data preprocessing, feature engineering, and machine learning while deriving actionable insights to improve customer retention.

Objective:
Participants will work with the **Telco Customer Churn Dataset** to:

1. Build a KNIME workflow for predicting customer churn.
2. Analyze the key factors contributing to customer churn using visualization and feature importance nodes.
3. Propose actionable recommendations to improve customer retention.

Dataset Details:

- **Source:** Telco Customer Churn Dataset on Kaggle
- **Size:** ~7,000 rows
- **Features:**
    - **Customer Demographics:**
        - gender: Gender of the customer.
        - Senior Citizen: Whether the customer is a senior citizen (1: Yes, 0: No).
        - Partner: Whether the customer has a partner (Yes/No).
        - Dependents: Whether the customer has dependents (Yes/No).
    - **Service Information:**
        - PhoneService, MultipleLines: Details of phone services.

        - InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies: Details of internet and other services.
    - **Financial Information:**
        - MonthlyCharges: Monthly charges incurred by the customer.
        - TotalCharges: Total amount charged to the customer.
    - **Contract Information:**
        - Contract: Type of contract (Month-to-month, One year, Two year).
        - PaymentMethod: Method of payment (e.g., Electronic check, Credit card).
        - Churn: Target variable indicating if the customer churned (Yes/No).

Challenge Tasks:

1. **Building the KNIME Workflow (20 points):**
    - Import the dataset into KNIME using the **File Reader** or **Excel Reader** node.
    - Visualize the dataset with **Data Explorer** and **Statistics** nodes.
    - Use **Missing Value** and **Column Filter** nodes to handle missing or irrelevant data.
2. **Data Preprocessing (20 points):**
    - Use **One-to-Many** nodes to encode categorical variables (e.g., Contract, InternetService).
    - Normalize numerical variables (e.g., MonthlyCharges, TotalCharges) with **Normalizer** nodes.
    - Apply the **Math Formula** node to create new features, such as tenure-to-total charges ratio.
3. **Exploratory Data Analysis (EDA) (20 points):**
    - Create visualizations using **Bar Chart**, **Pie Chart**, and **Box Plot** nodes to analyze key features.
    - Use the **Color Manager** node to visualize churn trends by customer groups.
    - Perform correlation analysis using the **Correlation Matrix** node.

4. **Predictive Modeling (30 points):**
    - Split the dataset into training and testing sets using the **Partitioning** node.
    - Build machine learning models using:
        - **Logistic Regression Learner**
        - **Random Forest Learner**
        - **XGBoost Tree Ensemble Learner**
    - Evaluate models with **Scorer**, **ROC Curve**, and **Confusion Matrix** nodes.
    - Compare model performance and select the best model based on accuracy, precision, recall, and F1-score.
5. **Insights and Recommendations (10 points):**
    - Identify the top factors influencing churn using the **Feature Importance** node.
    - Create visualizations with **Line Plot** and **Histogram** nodes to explain patterns in churn behavior.
    - Provide actionable business recommendations based on insights derived from the model and analysis.

Rubric:

Category	Points	Details
KNIME Workflow Design	20	Logical flow and proper use of KNIME nodes for data loading, preprocessing, modeling, and visualization.
Data Preprocessing	20	Effective handling of missing data, encoding, normalization, and feature engineering.
Exploratory Data Analysis (EDA)	20	Quality of visualizations and depth of analysis using appropriate KNIME nodes.
Predictive Modeling	30	Model accuracy, proper evaluation, and selection of the best-performing model.
Insights & Recommendations	10	Quality of insights derived and practicality of recommendations to reduce churn.

Expected Outcomes:

1. A KNIME workflow demonstrates end-to-end analysis and modeling.
2. Accurate predictions of customer churn.
3. Visualizations highlighting key drivers of churn.
4. A detailed summary report with actionable recommendations for improving customer retention.

Evaluation Criteria:

- Logical and efficient design of the KNIME workflow.
- Effective use of KNIME nodes for preprocessing, modeling, and visualization.
- Robustness and interpretability of the predictive model.
- Depth of insights and clarity of final recommendations.

Instruction Guidelines

1. **KNIME Workflow Design**:
    - Your workflow should be organized and logically structured, ensuring a seamless flow from data import to preprocessing, modeling, and visualization.
    - Properly label and document your nodes for clarity and ease of understanding. This is crucial for the evaluation of the workflow.
2. **Data Preprocessing**:
    - Handle missing or irrelevant data using the appropriate KNIME nodes such as Missing Value or Column Filter.
3. **Exploratory Data Analysis (EDA)**:

- Visualizations should be informative and help uncover key trends. Use appropriate visualization nodes like Bar Chart, Pie Chart, Box Plot, or Correlation Matrix to analyze data and churn patterns.

4. **Predictive Modeling**:
    - Use the appropriate learning algorithms: Logistic Regression, Random Forest, or XGBoost to model the data.
    - Ensure proper evaluation of your models using performance metrics such as accuracy, precision, recall, F1-score, ROC Curve, and Confusion Matrix.
    - Compare models and select the best-performing one based on the evaluation criteria.
5. **Insights and Recommendations**:
    - Create visualizations that clearly show churn trends and customer behavior patterns. Use these insights to provide actionable recommendations to improve customer retention.
6. **Collaboration and Teamwork**:
    - This is a group Hackathon, so effective collaboration is essential. Communicate regularly with your team members and ensure that all parts of the workflow are covered.
    - Make sure to check your work for accuracy and consistency. A team leader should be designated to oversee the overall quality of the project.

Deliverables

1. **KNIME Workflow**:
    - Submit a well-organized KNIME workflow that performs all the tasks outlined in the challenge (data import, preprocessing, modeling, and analysis). The workflow should be uploaded to the designated shared folder.
2. **Final Report (PowerPoint Presentation)**:
    - Create a PowerPoint presentation that summarizes your work, findings, and recommendations. The presentation should include:
        - An overview of the process and methodology used.
        - Key findings from the data preprocessing and exploratory analysis.

        - Visualizations highlighting churn patterns and key drivers.
        - Evaluation of the models used and the rationale for selecting the best model.
        - Actionable recommendations to reduce churn and improve customer retention.
    - The presentation should **not exceed 10 slides**.
    - Upload your presentation to the designated shared folder.
3. **Folder Structure**:
    - Please create **two separate folders** for your submission:
        - **a) Imported Workflow**: This folder should contain the final KNIME workflow.
        - **b) Final Report**: This folder should contain the PowerPoint presentation.
4. **Deadline**:
    - All submissions must be uploaded to the shared folders by **April 13th, 10:00 PM**. The folders will be **locked** after the deadline, and no revisions or resubmissions will be allowed.

5. **Resources:**

- Learning Center: https://knime.com/learning
    - Instructor-led Online Courses
    - Self-paced Courses
- Videos – KNIME TV YouTube Channel: youtube.com/user/KNIMETV
- Cheat Sheets: https://knime.com/cheat-sheets
- KNIME Press: https://knime.com/knimepress