Predicting Call Trends and Customer Churn Using Weather Data

An Analysis of Telecom Churn and Weather Patterns in the United States Using K-Nearest Neighbors (KNN)

Presented By: Donn Bryan Julian

Agenda

- Data Sources
- ETL Process
- Exploratory Data Analysis (EDA)
- Modeling
- Model Results (KNN)
- Key Findings
- Recommendations
- Next Steps

Data Sources

- **Telecom Churn Dataset**:

Source: Telecom Churn Clean Dataset on Kaggle

Description: Contains customer data, churn status, area codes, call usage details.

- **Weather Dataset**:

Source: Global Weather Repository on Kaggle

Description: Historical weather data by country and region (temperature, precipitation, humidity).

ETL Process

- **Extract**:

Pulled the **telecom churn dataset** and **weather dataset** from Kaggle.

Focused on **California** area codes for granular analysis.

- **Transform**:

Cleaned and filtered both datasets.

Merged weather data with telecom data using **area codes.**

Removed irrelevant columns (e.g., non-U.S. data).

- **Load**:

Combined dataset used for **training and evaluation** in modeling phase.

Exploratory Data Analysis (EDA)

- Objective of EDA:

Explore the relationship between weather data and customer churn.

Identify the key features influencing telecom service usage and churn.

- Insights Derived:

Customer Service Calls:

  - There's a noticeable correlation between customer service calls and churn (0.21).

Weather-Related Features:

  - Variables like temperature (0.02) and precipitation show minor correlations with churn, indicating more subtle patterns, likely better captured by non-linear models like KNN.

- Key Visualizations:

Correlation Heatmap:

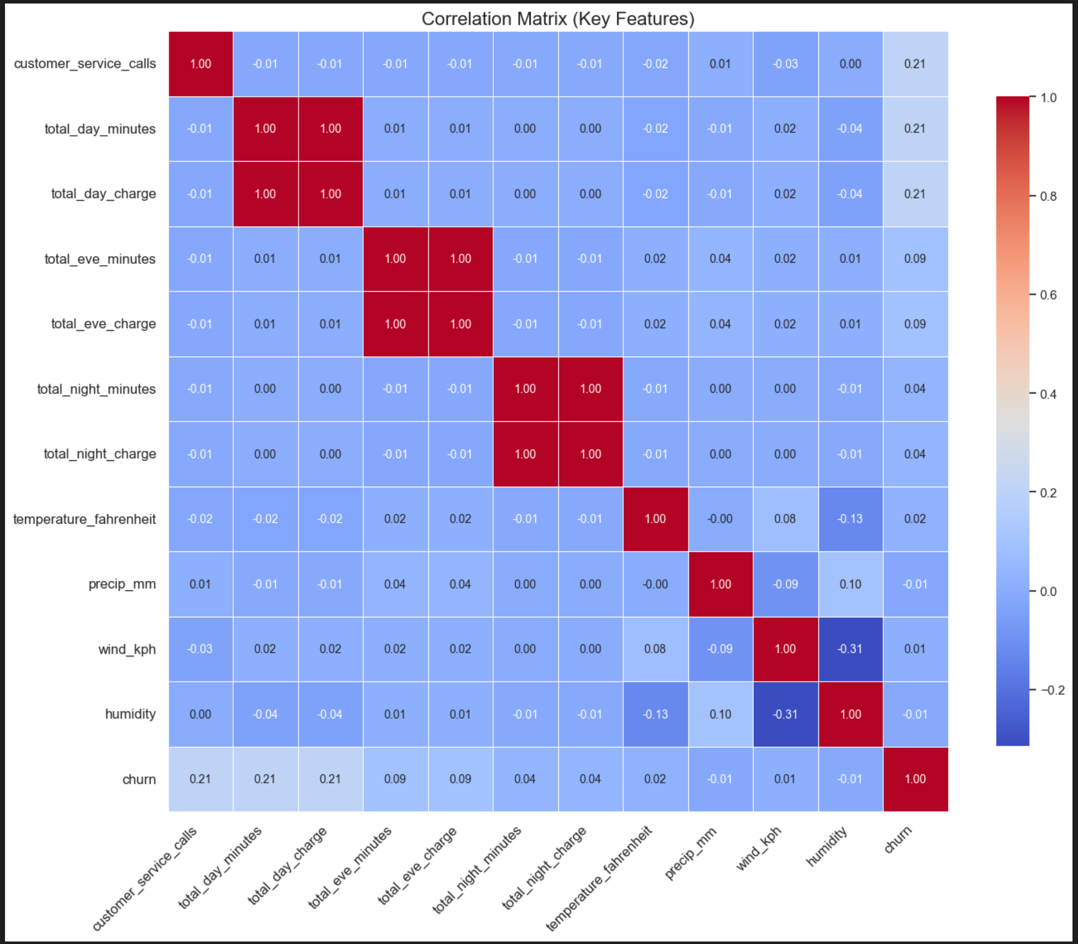
  - The matrix highlights relationships between key features such as call behavior, weather patterns, and churn.
  - Strong correlations between telecom usage metrics (e.g., day minutes and day charges), while churn is moderately correlated with customer service calls.
  - *Place the correlation matrix image here* (Resize it if needed to fit well within the slide layout).

- Feature Selection:

The EDA led to the selection of:

  - Key Telecom Features: Customer service calls, total day minutes, and charges.
  - Weather Features: Temperature, precipitation, wind speed, and humidity.

Customer service calls and telecom usage metrics show moderate correlations with churn. Weather variables have weaker linear correlations, suggesting possible non-linear relationships that models like KNN can capture.



Modeling Process (KNN)

Brief Description:

For this analysis, we selected K-Nearest Neighbors (KNN) to capture potential non-linear relationships between weather conditions and churn.

The confusion matrix shows that the model achieved high accuracy in predicting both churned and non-churned customers, with only a small number of misclassifications.

True Positives (1,1): 78 correctly predicted churns.

True Negatives (0,0): 573 correctly predicted non-churns.

Only 16 misclassifications (1 false positive and 15 false negatives).

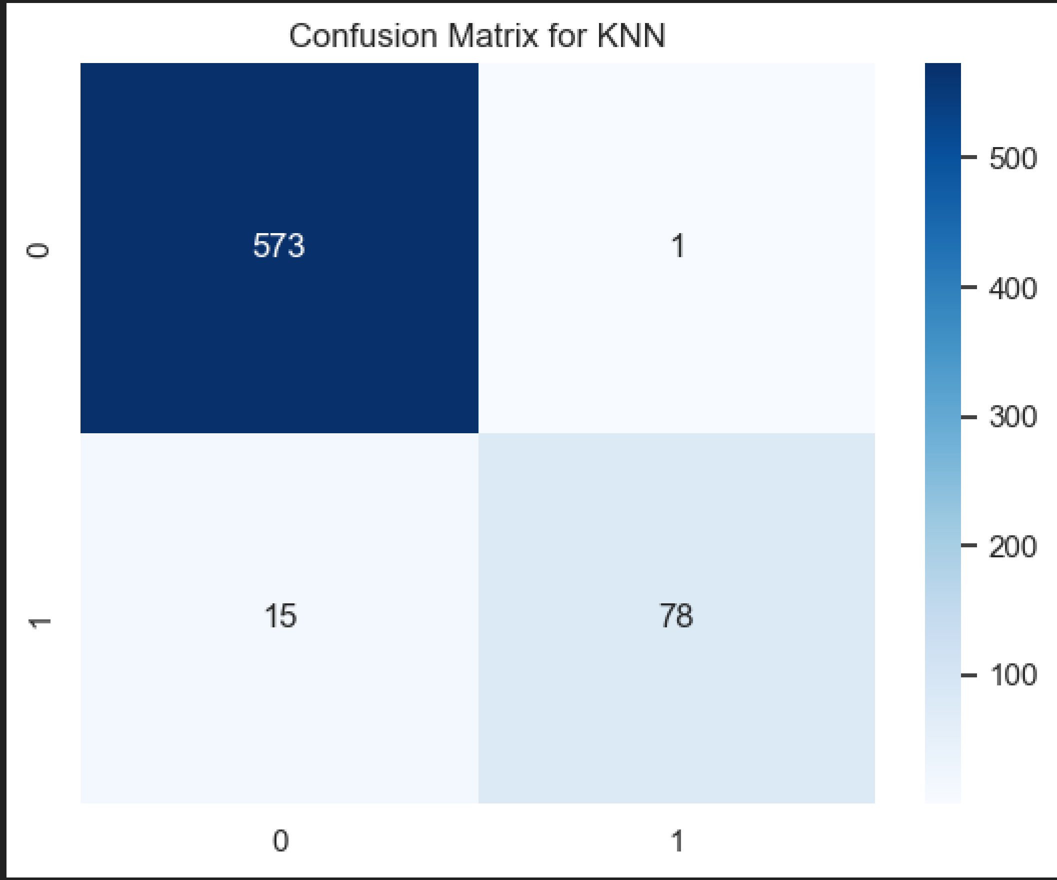
Classification Report (Text or a simple table):

Accuracy: 98%

Precision for class 0 (Non-Churn): 0.97, class 1 (Churn): 0.99

Recall for class 0: 1.00, class 1: 0.84

F1-score: 0.99 for class 0, 0.91 for class 1



Key Findings

1. **Churn is correlated with customer service calls**:

   The model confirms that **frequent customer service interactions** are a strong predictor of churn.

   Customers who made multiple service calls were more likely to churn.

2. **Telecom Usage Metrics impact churn**:

   High usage during the day (total day minutes and charges) has a moderate correlation with churn, likely due to **bill shock** or service dissatisfaction.

3. **Weather conditions have weak linear correlations**:

   Weather features like **temperature** and **precipitation** showed weak direct correlations with churn, but the **KNN model** helped reveal non-linear relationships.

4. **KNN Model Performance**:

   The model achieved a high **98% accuracy** with strong **precision** and **recall** for both churned and non-churned customers.

   The model performed well at predicting churn, with a **0.99 precision** for churned customers and an **F1-score** of **0.91**.

Recommendations

Recommendations:

1. **Proactive Customer Support for High-Risk Users**:

   Increase customer support resources for customers making frequent service calls, as they are more likely to churn.

   Implement **targeted retention strategies** for customers identified as high-risk by the model (e.g., those with high usage and frequent service calls).

2. **Weather-Based Campaigns**:

   Use **weather data** to anticipate service issues or potential dissatisfaction in areas experiencing **extreme weather conditions**.

   Offer **incentives** or **discounts** to customers in regions affected by severe weather to reduce churn.

3. **Refine Telecom Packages Based on Usage**:

   Analyze **high usage patterns** (e.g., total day minutes and charges) to optimize telecom packages, avoiding **bill shock** that could drive churn.

   Consider creating **customized plans** for heavy users to enhance customer satisfaction and retention.

4. **Deploy KNN Model for Real-Time Churn Prediction**:

   Integrate the **KNN model** into your systems for real-time monitoring of churn risk.

   Use the model's insights to predict churn and adjust customer service efforts dynamically.

Next Steps

• **Deploy the KNN Model**:

Integrate the trained **KNN model** into the telecom's existing systems for **real-time churn prediction**.

Ensure that customer service teams and marketing departments have access to these predictions for proactive outreach.

• **Monitor and Optimize Model Performance**:

Set up **regular performance reviews** for the model to ensure it maintains high accuracy as new data is introduced.

Continuously retrain the model with **fresh data** to capture evolving customer behavior and changing weather patterns.

• **Expand the Model**:

Extend the current model beyond California to other regions, incorporating more **geographic data** to predict churn across the country.

Add other **external factors** (e.g., economic data, service outages) to improve prediction accuracy and cover more scenarios.

• **Improve Customer Experience Based on Insights**:

Use the churn predictions to **optimize customer service** workflows, increasing support in high-risk areas during severe weather events.

Personalize marketing campaigns based on usage patterns and churn risk levels to **increase retention**.

Q&A

Thank you for your attention.
We're happy to take any questions you may have