

Faculty of Management Technology
Artificial Intelligence (ICSEN 933)

Assignment 2 Report

Submitted by:

| | | |
|-------------------------|-------------|-------------|
| Name: Ali Refaat Geady | ID: 58-4020 | Tutorial: 7 |
| Name: Ahmed Hassan | ID: 58-0671 | Tutorial: 7 |
| Name: Omar Sherif Kadry | ID: 58-1335 | Tutorial: 7 |
| Name: Ziad Ekramy | ID: 58-6936 | Tutorial: 7 |

Supervised by:

Dr. Mohamed Karam Gabr

TA Mahinar Reda

Submission Date:

30/11/2025

Customer churn prediction is a problem because businesses in the telecom sector are always at risk of losing clients to competitor service providers. Churn occurs when customers cancel their subscriptions and switch to another business. By anticipating this behavior, service providers can take early action with focused marketing tactics. Since it usually costs much more to acquire new consumers than to keep old ones, this issue is crucial. Thus, a company's revenue, customer satisfaction, and long term competitiveness are directly and significantly impacted by precisely identifying consumers who are likely to leave. The ability to predict churn offers a significant strategic advantage in an industry as competitive as telecommunications. Despite its importance, churn prediction presents several challenges. The Telco-Customer dataset, which is used to solve this issue, is inherently unbalanced since the majority of consumers do not leave. Traditional models may be biased toward predicting the majority class because only roughly 27% of consumers fall into the turnover class. That's why the Support Vector Machine (SVM) is a perfect fit. In this assignment, the methodology involved analyzing the dataset from Kaggle, preprocessing the data to ensure quality, and turning features into numerical columns to be processed by the SVM. For training and testing, we used an 80/20 split then compared the actual and the predicted result. This method helps us figure out the chances of customers changing their behavior throughout their relationship with a telecommunication company. It gives us an idea of how likely customers will churn as time goes on, so we can create smart, data backed plans to keep them.

The Telco Customer Churn dataset was chosen because it offers a thorough and accurate depiction of consumer behavior in the telecom sector. It contains a wide range of variables that are essential for assessing client engagement and forecasting churn patterns, including demographic information, account tenure, contract type, monthly charges, and churn status. This dataset was also chosen because of its clean, organized nature, which facilitates quick preprocessing and effective feature engineering. Additionally, the dataset's clarity makes it perfect for showing how customer lifecycle stages change over time, offering analytical depth and obvious business significance. Additionally, the data science community has publicly acknowledged and validated this dataset, guaranteeing its dependability and comparability with other churn Prediction studies. It has received 445 thousand downloads and 2.58 million views

overall. Lastly, its real-world setting enables significant insights into client retention tactics, which is exactly in line with our objective of using an SVM to analyze and comprehend churn behavior.

The accuracy and dependability of customer churn prediction are significantly influenced by the characteristics used for the model. Together, these characteristics offer a thorough picture of consumer behavior, which makes them crucial. A customer's level of engagement and financial commitment to the service provider are shown in attributes like tenure, MonthlyCharges, and TotalCharges, which have a significant impact. Larger monthly fees may make some client segments unhappy, whereas shorter tenure frequently suggests a larger chance of churn. InternetService, TechSupport, OnlineSecurity, and Contract type are examples of service attributes that are also important predictors since consumers who are locked into long-term contracts or who rely substantially on services are less likely to leave. Additional categorical variables like Partner, Dependents, and PaymentMethod complement the behavioral data by capturing demographic and preference driven influences on retention. Together, these encoded features form a multidimensional profile that allows the Support Vector Machine model to identify meaningful patterns associated with customer churn.

To guarantee that the Support Vector Machine model could function efficiently and that the input characteristics appropriately reflected the underlying structure of the data, preprocessing techniques were chosen for the Telco-Customer dataset. Cleaning the TotalCharges column, where non-numeric values were found and eliminated, was the first crucial preprocessing step. Retaining rows with incorrect data would create mistakes and skew the model's understanding of customer spending behavior because SVMs only accept numerical inputs. Z-score normalization was then used to standardize numerical parameters including tenure, MonthlyCharges, and TotalCharges. Because SVM optimization relies on distance calculations, this scaling was required. Without normalization, the model will output biased predictions. Standardization guarantees that every numerical characteristic makes an equal contribution to the model. One-hot encoding was used to encode categorical information, such as gender, InternetService, Contract, PaymentMethod, and several service indications. One-hot encoding was selected since there is no

hierarchy order to categorical values. In order to guarantee that all input attributes were numerical and compatible with the SVM, boolean features like Partner, Dependents, PhoneService, and service subscription indicators, and more columns were also encoded into binary format. Finally, the target characteristic, Churn, was translated from "Yes/No" to 1 and 0 before being transformed into +1 and -1. The dataset was then split 80/20 for training and testing.

As a result of testing, our SVM model achieved an accuracy of approximately 81%, meaning that it correctly classified the majority of churn outcomes. Considering the significant class imbalance in the dataset, where about 73% of customers did not churn and only 27% did, this level of performance indicates that the model successfully captured meaningful patterns rather than simply predicting the majority class. Furthermore, the comparison table between the predicted and actual labels demonstrates strong alignment across most instances, reinforcing the reliability of the model. Overall, the SVM implementation can be considered both successful and satisfactory.

When comparing the results of the Markov Model with the results from the SVM model, a clear distinction emerges in how each approach handles customer churn prediction. The Markov Model focuses on estimating transition probabilities between customer states over time, revealing that "Churned" consistently appears as one of the top three most probable future states across nearly all customer segments. For example, customers in New_Month-to-month have a 26.6% chance of transitioning to Churned, while Active_Month-to-month customers show an even higher 28.5% probability of churning. This pattern reflects the Markov Model's strength in identifying long term behavioral trends and highlighting customer groups that possess inherently higher churn risks. However, despite its usefulness in describing state dynamics, the Markov Model is limited in predictivity because it assumes future behavior depends only on the current state and not on the full set of customer attributes. In contrast, the SVM model integrates numerous customer features, tenure, monthly charges, contract type, and service usage, allowing it to capture more complex, non-linear relationships that influence churn. The SVM achieved an

accuracy of approximately 81%, outperforming the Markov Model's transitions by offering a more actionable, instance prediction. Unlike the Markov results, which often exaggerated churn as a dominant transition destination, the SVM provides a more accurate classification based on real attribute patterns. This makes the SVM far more effective in practical churn prediction, as it not only identifies high risk groups but also differentiates individual customers with a higher degree of reliability. Overall, while the Markov approach is valuable for understanding customer lifecycle movements and general behavioral trends, the SVM model demonstrates superior predictive capability and produces results that are ultimately more satisfactory for operational decision making.

Dataset: [Telco Customer Churn](#)

Copy of Notebook (Source Code): [!\[\]\(2e897e890e69d81eae4503a8342c36b0_img.jpg\) Assignment 2 Implementation.pdf](#)