



CHURN PREDICTION USING AI-ML

Name : Yojan Gandhi
Roll Number: 33229
Class : TE-10

Guide

Mrs. Sumitra Jakhete Maam

Reviewer

Mr. Abhijeet Karve Sir

PURPOSE :

Churn prediction is a crucial tool for businesses, acting like a crystal ball that helps them foresee which customers may stop using their products or services. This is done by analyzing customer behavior and patterns using data and smart algorithms. Once at-risk customers are identified, businesses can take proactive steps to retain them, such as offering discounts or improved service. The ultimate aim is to prevent customer departures, as losing customers can negatively impact a company's revenue and reputation while retaining them leads to increased loyalty and profitability.

Literature Survey:

Title	Authors	Publication	Summary
A Comprehensive Survey of Customer Churn Prediction Techniques in the Telecommunication Sector	Hafeez Anwar, Habin Lee, Hong S. Yoon	IEEE Access, 2019	In-depth analysis of churn prediction techniques, methodologies, datasets, and performance metrics in the telecommunications sector.
Churn Prediction in E-commerce: A Review and Future Directions	Sameera Mubarak Ali, Helen Ashman	Expert Systems with Applications, 2021	Comprehensive review of churn prediction techniques in e-commerce. Provides insights into future research directions in the field.
Churn Prediction in the Gaming Industry: A Comprehensive Survey	Jorge M. Oliveira, et al.	Expert Systems with Applications, 2021	Explores approaches and methodologies for predicting churn among players in the gaming industry.
Artificial IntelligenceBasedCustomerChurn PredictionModel for Business Markets	J. Faritha Banu ,1 S. Neelakandan,2 B.T Geetha,3 V. Selvalakshmi,4 A.Umadevi	2022	Application of various machine learning techniques to predict customer churn in the telecom industry, emphasizing the importance of model interpretability and data-driven retention strategies.

SCOPE :

The scope of this churn prediction project encompasses collecting and analyzing relevant customer data, developing predictive models to forecast customer churn, deploying these models within the business operations for real-time predictions, and implementing actionable strategies to retain at-risk customers. Continuous monitoring, maintenance, and documentation of the project's processes and outcomes are essential components of its scope, with a focus on improving customer retention, enhancing profitability, and ensuring ethical data handling throughout the project lifecycle.

OBJECTIVE :

The primary objective of churn prediction is to leverage data analytics and predictive modeling to anticipate which customers are likely to discontinue their engagement with a business or cease using its services. By proactively identifying these at-risk customers, companies can initiate targeted strategies aimed at retaining them, such as personalized offers, improved customer support, or product enhancements. Beyond reducing customer attrition and preserving revenue, the broader goals encompass enhancing overall customer satisfaction and loyalty. Churn prediction also facilitates data-driven decision-making within organizations, optimizing resource allocation for customer retention efforts and fostering a competitive advantage by demonstrating a commitment to customer-centric practices. In essence, it's a strategic approach that contributes not only to short-term financial stability but also to long-term business growth and resilience.



ABSTRACT



Recent advancements in technology, specifically artificial intelligence (AI) and machine learning (ML), have brought exciting changes to how businesses work better. One big challenge for companies, especially those in telecommunications, is predicting when customers might decide to leave their services. This can cause problems for a company's money situation. To tackle this, companies are using AI and ML to figure out beforehand if customers might leave. This study has two main parts: first, a detailed plan is made to predict when customers might leave. This involves getting data ready, looking at important parts, and picking the best pieces of data. Then, the data is split into two groups: one for learning and one for testing how well the plan works. Different computer models are used to guess if customers will leave or stay. Some models work better when they work together. The second part introduces a new way to make these guesses. Smart algorithms help pick out important things that tell us if someone is leaving. These algorithms also help decide if someone is leaving or not. The study shows that using AI and ML can help predict when customers might leave a company. This helps companies make their customers happier and make more money. It shows how to make good predictions that work well even with different information.



Keywords: advancements, technology, artificial intelligence, AI, machine learning, ML, businesses, telecommunications, predicting, services, money, AI and ML, computer models, algorithms, decisions, predictions, information.

What is Churn ?

Churn refers to the phenomenon where customers or users stop using a company's products or services. It's also known as customer attrition or customer turnover.

Customers leaving a particular companies service or stop using its product.



BASIC TERMINOLOGY OF CHURN

- **Churn:** Churn refers to the phenomenon where customers or users stop using a company's products or services. It's also known as customer attrition or customer turnover.
- **Churn Rate:** The churn rate is the percentage of customers who have churned during a specific period, usually measured monthly or annually. It's a key metric for assessing customer retention.
- **Churn Prediction:** Churn prediction is the process of using data and predictive analytics to forecast which customers are likely to churn in the future. It involves building models to identify at-risk customers.
- **Customer Lifecycle:** The customer lifecycle is the journey that a customer goes through from the initial engagement with a company (acquisition) to becoming a loyal, long-term customer (retention) or potentially churning.
- **Retention:** Retention refers to the ability of a company to keep its existing customers engaged and continue using its products or services over time.
- **Customer Segmentation:** Customer segmentation involves dividing customers into groups or segments based on characteristics such as behavior, demographics, or purchase history. Segmentation helps tailor retention strategies.
- **Feature Engineering:** Feature engineering is the process of selecting and creating relevant features (variables) from customer data that can be used in churn prediction models. These features help identify churn patterns.
- **Predictive Modeling:** Predictive modeling uses statistical and machine learning techniques to build algorithms that can predict future outcomes, such as customer churn. Common models include logistic regression, decision trees, and neural networks.
- **Training Data:** Training data is the historical dataset used to train a churn prediction model. It includes past customer behavior and whether they churned or not.

BASIC TERMINOLOGY OF CHURN

- **Test Data:** Test data is a separate dataset used to evaluate the performance of a churn prediction model. It allows for assessing how well the model generalizes to new, unseen data.
- **False Positives and False Negatives:** In churn prediction, a false positive occurs when the model predicts a customer will churn, but they do not. A false negative occurs when the model predicts a customer will not churn, but they do. Balancing these errors is crucial for model accuracy.
- **ROC Curve and AUC:** The Receiver Operating Characteristic (ROC) curve is a graphical representation of a model's performance. The Area Under the Curve (AUC) is a summary metric that measures the model's ability to distinguish between churn and non-churn cases. **Threshold:** The threshold is a value used to determine the cutoff point for classifying customers as potential churners or non-churners based on model predictions. Adjusting the threshold can trade off between false positives and false negatives.

CONCEPTS OF CHURN

- **Data Collection:** Collect and store relevant customer data, including transaction history, user interactions, demographics, and customer feedback. High-quality data is essential for accurate churn prediction.
- **Feature Engineering:** Create meaningful features or variables from the collected data that can serve as inputs for churn prediction models. These features can include customer behavior metrics, customer lifetime value, and engagement scores.
- **Labeling Churn:** Determine how to label customers as churners or non-churners in your historical dataset. This labeling process is crucial for training supervised machine learning models.
- **Training Data:** Prepare a historical dataset with labeled churn outcomes (churned or not churned) to train machine learning models. This dataset should cover various customer behaviors and demographics.
- **Test Data:** Set aside a separate dataset for model evaluation and validation. This data should not be used during model training to assess how well the model generalizes to new data.
- **Supervised Learning:** Churn prediction typically falls under supervised learning, where models learn from historical data with known outcomes to make predictions about new, unseen data. Predictive Models: Experiment with different machine learning algorithms such as logistic regression, decision trees, random forests, support vector machines, and neural networks to build churn prediction models.
- **Evaluation Metrics:** Use appropriate evaluation metrics like accuracy, precision, recall, F1-score, and ROC AUC to assess the performance of churn prediction models. The choice of metric depends on the specific business goals and the cost of false positives and false negatives.

CONCEPTS OF CHURN

- **Imbalanced Data:** Address class imbalance if the dataset has significantly more non-churners than churners. Techniques like oversampling, under sampling, or using different sampling strategies can help balance the dataset.
- **Feature Importance:** Understand which features have the most significant impact on predicting churn. Feature importance analysis helps prioritize retention efforts.
- **Hyperparameter Tuning:** Fine-tune the hyperparameters of machine learning models to optimize their performance. Grid search or random search can be used for this purpose.
- **Threshold Selection:** Decide on an appropriate threshold for model predictions to classify customers as potential churners. Adjusting the threshold can impact the balance between false positives and false negatives.
- **Cross-Validation:** Implement cross-validation techniques to assess model robustness and avoid overfitting. Common methods include k-fold cross-validation.
- **Deployment:** Deploy the trained churn prediction model into the business's operational environment, allowing for real-time predictions or periodic batch predictions.
- **Monitoring and Maintenance:** Continuously monitor the model's performance, retrain it periodically with new data, and update it as customer behavior patterns change.
- **Actionable Insights:** Translate model predictions into actionable strategies for customer retention, such as targeted marketing campaigns, personalized offers, or customer service improvements.

TECHNIQUES EXPLORED

- **Machine Learning Algorithms:**

- **Logistic Regression:** A straightforward algorithm for binary classification.
- **Decision Trees and Random Forests:** Effective for capturing complex decision boundaries.
- **Support Vector Machines (SVM):** Useful for separating data into different classes.
- **Neural Networks:** Deep learning models can capture intricate patterns in large datasets.

- **Ensemble Methods:**

- **Boosting (e.g., AdaBoost, Gradient Boosting):** Combines multiple weak models to create a stronger predictive model.
- **Bagging (Bootstrap Aggregating):** Reduces variance by training multiple models on bootstrapped subsets of the data.

- **Time-Series Analysis:**

- Examine time-dependent patterns in customer behavior to predict future churn.
- Techniques like ARIMA, seasonal decomposition, and Prophet can be useful.

- **Survival Analysis:**

- Especially relevant when analyzing customer churn over time, considering factors like time to churn and censoring.
- Kaplan-Meier estimation, Cox Proportional-Hazards models, and parametric survival models are commonly used.

- **Clustering and Segmentation:**

- Cluster customers into groups with similar behaviors or characteristics.
- K-Means, hierarchical clustering, and DBSCAN can help create customer segments.

- **Natural Language Processing (NLP):**

- If textual data, such as customer feedback or comments, is available, NLP techniques can extract valuable insights.
- Sentiment analysis, topic modeling, and text classification can be employed.

- **Feature Engineering:**

- Create informative features, such as customer lifetime value, recency-frequency-monetary (RFM) metrics, and customer engagement scores.
- Use domain knowledge to engineer features that are relevant to the business context.

TECHNIQUES EXPLORED

- **Hyperparameter Tuning:**
 - Fine-tune the hyperparameters of machine learning models to optimize their performance.
 - Grid search or random search can be used for this purpose.
- **Threshold Selection:**
 - Decide on an appropriate threshold for model predictions to classify customers as potential churners.
 - Adjusting the threshold can impact the balance between false positives and false negatives.
- **Cross-Validation:**
 - Implement cross-validation techniques to assess model robustness and avoid overfitting.
 - Common methods include k-fold cross-validation.
- **Deployment:**
 - Deploy the trained churn prediction model into the business's operational environment, allowing for real-time predictions or periodic batch predictions.
- **Monitoring and Maintenance:**
 - Continuously monitor the model's performance, retrain it periodically with new data, and update it as customer behavior patterns change.
- **Actionable Insights:**
 - Translate model predictions into actionable strategies for customer retention, such as targeted marketing campaigns, personalized offers, or customer service improvements.

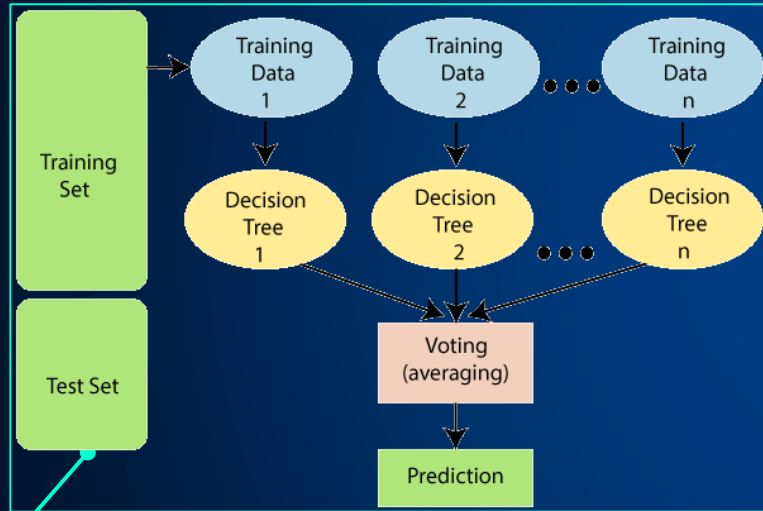
Summarizing :

- Dataset can be used : Churn.csv

- Data Pre-processing is required as this removes the error while model prediction takes place.
- Then comes the classification where we need to split our dataset into training and testing.
- Train your model using supervised learning to ensure high accuracy of the model.

- Apply decision tree classification on the various parameters of the models.
- Now use each of the decision tree outcome to generate a random forest output.
- Test your model on different datasets.

Random Forest Use in Churn Prediction Model

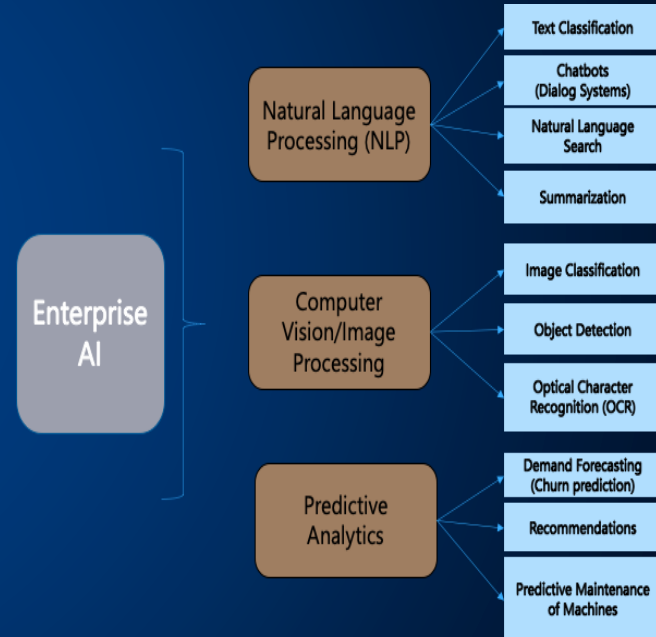


- Random Forest is a powerful ensemble learning technique that overcomes the limitations of single decision trees by constructing a forest of randomized decision trees. Through a process of bagging and feature selection, it improves model accuracy and robustness, making it a valuable tool in various domains. Random Forest's ability to handle large feature sets, assess feature importance, and provide robust predictions makes it particularly suitable for applications like customer churn prediction, fraud detection, and more. By understanding its key characteristics and best practices, businesses can harness the potential of Random Forest for enhanced predictive modeling.

Random forest is better than decision tree as when we are using outputs of multiple decision trees we reduce the over fitting of the curve and hence improving the over overall accuracy of the model

NLP(Natural Language Processing)

- Natural Language Processing (NLP) is used in churn prediction to extract valuable insights from textual data, such as customer reviews, feedback, or communication logs.
- NLP in churn prediction leverages text analysis techniques to process and analyze unstructured textual data, allowing businesses to gain a deeper understanding of customer sentiments and feedback. By extracting key information from sources like customer reviews or support chat logs, NLP helps identify early warning signs of potential churn. Sentiment analysis, topic modeling, and text classification are common NLP methods applied to assess customer satisfaction, discover underlying issues, and inform retention strategies. This enables companies to proactively address customer concerns and reduce churn by improving their products, services, or support.



A quick comparison in Gaming Industry

- Appeal of New Elements
- Desire for Variety
- Community and Friends
- Competitive Scene



CHURN



APPLICATIONS

Q1 TELECOMUNICATION

Implement targeted retention offers, service improvements.
Predict which mobile or internet subscribers are likely to switch to a different provider.

Q2 AUTOMOBILE INDUSTRY

Predict vehicle owners considering changing car brands.
Offer maintenance services, warranties, or loyalty incentives to retain customers.

Q3 HOSPITAL AND TRAVEL

Forecast travelers who may choose other destinations or accommodations.
Provide special offers, loyalty rewards, or personalized travel experiences to retain guests.

Q4 GAMING INDUSTRY

Forecast gamers who might stop playing or purchasing in-game items.
Offer personalized gaming content, discounts, or incentives to retain players.

APPLICATIONS

Q5 ONLINE MARKETPLACE

Identify sellers or buyers who may stop using the platform.

Provide support, advertising credits, or trust-building measures to retain users.

Q6 BANKING AND FINANCE

Forecast customers likely to close their accounts or switch banks.

Offer customized financial products, loyalty rewards, or better interest rates to retain customers.

Q7 NON-PROFIT ORGANIZATION

Identify donors or members who may cease contributing or participating.

Implement personalized engagement strategies and demonstrate the impact of their support to retain supporters.

Q8 INSURANCE

Predict policyholders who may not renew their insurance policies.

Tailor coverage options, adjust premiums, or provide additional benefits to retain policyholders.

APPLICATIONS

Q9 HEALTHCARE PROVIDERS

Identify patients at risk of switching healthcare providers. Improve patient care, appointment scheduling, and communication to enhance patient retention.

Q10 MANUFACTURING AND B2B SERVICES

Predict business clients considering switching suppliers. Offer volume discounts, customized solutions, or quality improvements to retain business partners.

Q11 FITNESS AND HEALTH APPS

Forecast users who may discontinue using fitness or wellness apps. Offer personalized workout plans, nutritional guidance, or challenges to retain users.

Q12 SUBSCRIPTION SERVICES[STREAMING]

Identify subscribers at risk of canceling their subscriptions. Personalize content recommendations, offer discounts, or extend trial periods to retain subscribers.

CONCLUSION :

In conclusion, the churn prediction project holds immense potential for businesses and organizations seeking to proactively manage customer attrition and enhance customer retention strategies. By leveraging data analytics, predictive modeling, and machine learning techniques, this project empowers decision-makers to anticipate which customers are likely to churn in the future, thereby allowing for timely and targeted interventions. Through the careful application of the techniques and concepts discussed, businesses can not only reduce customer churn but also improve customer satisfaction, optimize resource allocation, and gain a competitive edge in their respective industries .As businesses continue to prioritize data-driven approaches, the implementation of effective churn prediction systems remains a vital step toward achieving sustainable growth and resilience in an ever-evolving market landscape.

REFERENCES :

- J. Faritha Banu ,1 S. Neelakandan,2 B.T Geetha,3 V. Selvalakshmi,4 A. Umadevi,4and Eric Ofori Martinson Computational Intelligence and “Artificial IntelligenceBasedCustomerChurn PredictionModel for Business Markets” Neuroscience 5 Volume 2022, Article ID 1703696, 14 pages,Published 29 September 2022,<https://doi.org/10.1155/2022/1703696>
- I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," in IEEE Access, vol. 7, pp. 60134-60149, 2019,Published on May 6, <https://doi:0.1109/ACCESS.2019.2914999>
- 1.V. Kavitha,S. V Mohan Kumar,G. Hemanth Kumar,M. Harish,”Churn Prediction of Customer in Telecom Industry using Machine Learning Algorithms” Computational Intelligence and Neuroscience -IJERT -ISSN: 2278-0181,Published on 05, May-2020

CHURN PREDICTION USING AI AND ML

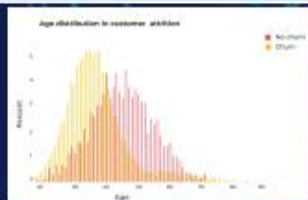
Yojan Gandhi : 33229

ABSTRACT

- **Tech Advancements:** AI and ML tech advances are changing how businesses operate.
- **Customer Churn Challenge:** Predicting when customers might leave, especially in telecom, is a major issue.
- **AI & ML Solutions:** Companies use AI and ML to predict and prevent customer departures, enhancing customer satisfaction and profits.

Churn Prediction Essentials

- **Data Collection & Preparation :** Gather customer data.
- **Clean and organize the data.**
- **Feature Selection & Engineering:** Identify key customer behavior indicators.
- **Model Selection & Training :** Choose a suitable prediction model.
- **Train the model on historical data.**
- **Deployment :** Implement the model for real-time predictions.
- **Monitoring & Feedback :** Continuously assess model performance.
- **Gather customer feedback for improvement.**
- **Iterate & Improve :** Refine the model and strategies over time.



Model Training

- **Data Preparation:** Clean and preprocess historical customer data.
- **Feature Engineering:** Extract relevant customer behavior features.
- **Data Splitting:** Divide data into training and testing sets.
- **Model Selection:** Choose an appropriate algorithm.
- **Model Training:** Train the selected model with the training data.
- **Cross-Validation:** Assess model performance and prevent overfitting.
- **Hyperparameter Tuning:** Fine-tune model settings for better accuracy.
- **Model Evaluation:** Measure accuracy using metrics like precision, recall, and ROC-AUC.
- **Deployment:** Implement the model for real-time predictions.
- **Continuous Monitoring:** Regularly check the model's performance.
- **Feedback Loop:** Gather customer feedback to refine strategies and enhance retention.

Decision Metrics

- **Accuracy:** Overall prediction correctness.
- **Precision:** Accuracy of positive churn predictions.
- **Recall (Sensitivity):** Correct identification of actual churn.
- **F1-Score:** Balance between precision and recall.
- **ROC-AUC:** Ability to distinguish churn from non-churn.
- **Confusion Matrix:** Detailed prediction breakdown.
- **Customer Lifetime Value (CLV):** Financial impact assessment.
- **Gini Coefficient:** Identifying high-risk customers.
- **K-Score:** Model effectiveness comparison.
- **MAE/RMSE:** Prediction accuracy measurement.



ARCHITECTURE



Ensemble Models in Churn Prediction Ensemble

Ensemble Models in Churn Prediction
Ensemble models, similar to the effectiveness of combining phrases in translation, amalgamate diverse predictive algorithms to enhance churn prediction accuracy. They offer a robust approach by considering multiple perspectives on customer behavior, resulting in more reliable churn forecasts. Ensemble methods like Bagging, Boosting, and Random Forest have demonstrated substantial success in improving churn prediction outcomes.

Reference

- J. Faritha Banu, I. S. Neelakandan, B. T. Geetha, V. Selvakumari, A. Umadevi, and Eric Offor Martinson Computational Intelligence and "Artificial Intelligence-Based Customer Churn Prediction Model for Business Markets" *Neuroscience 5 Volume 2022*, Article ID 1703696, 14 pages, Published 29 September 2022 [Add a little bit of body text](#)
- H. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," in *IEEE Access*, vol. 7, pp. 60134-60149, 2019, Published on May 6, <https://doi.org/10.1109/ACCESS.2019.2914999>
- V. Kavitha, S. V. Mohan Kumar, G. Hemant Kumar, M. Harish, "Churn Prediction of Customer in Telecom Industry using Machine Learning Algorithms" *Computational Intelligence and Neuroscience - IJERT* - ISSN: 2278-0181, Published on 05, May-2020

THANK YOU!!!

