# Customer Churn Prediction

- By Sameer Yadav (915), Aditya K. Singh (916), Aashish Kumar (905)

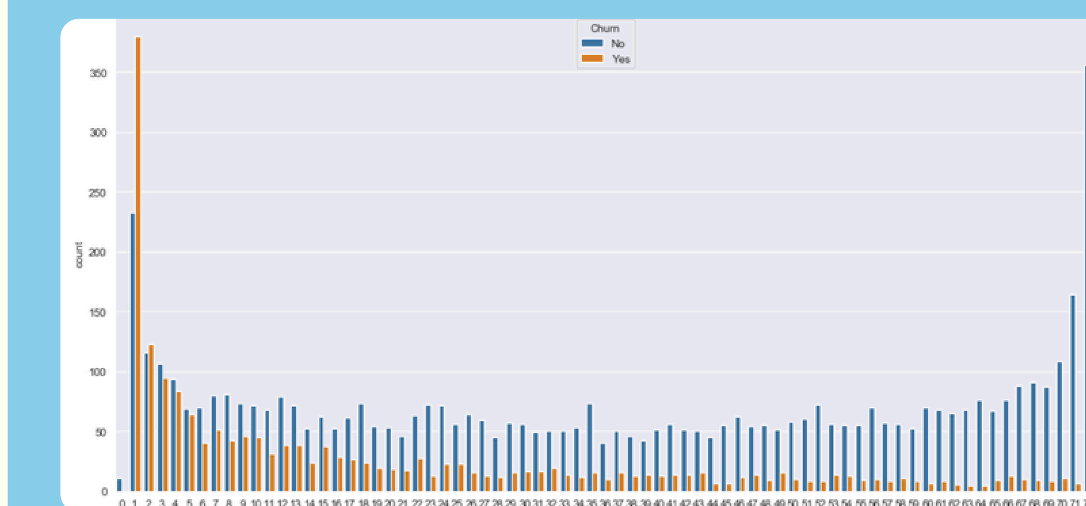
## One-line Description

Predicting customer churn and lifetime value using survival analysis and machine learning models.

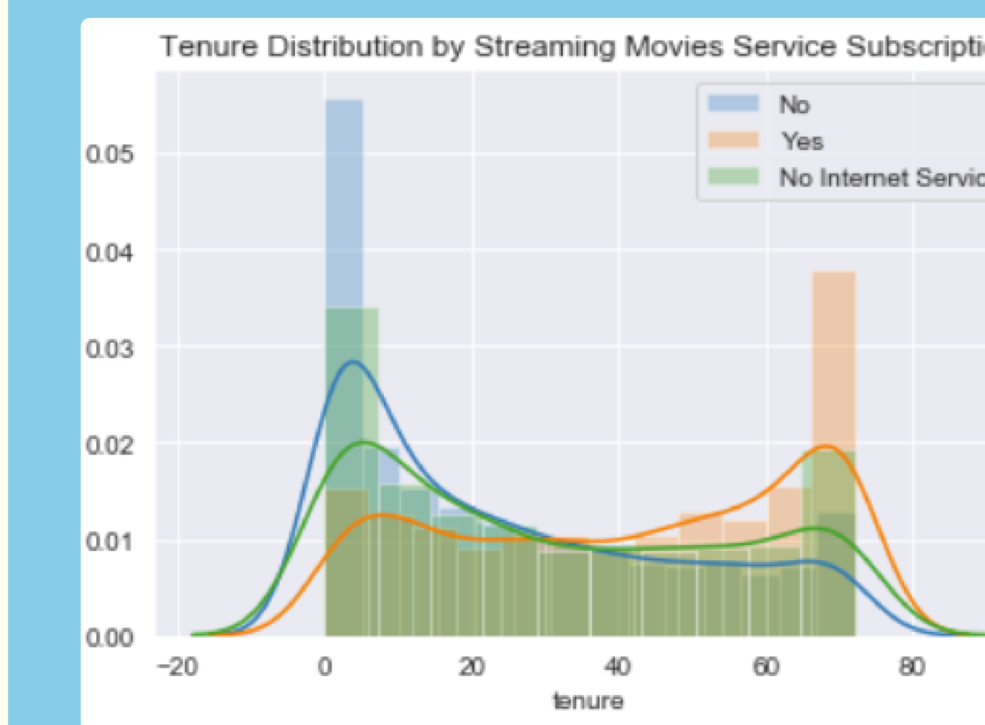
## Dataset Overview

- Dataset Name: Telco Customer Churn Dataset
- Source: Publicly available on Kaggle (or telecom company internal data)
- Total Records: 7043 customers
- Number of Columns: 21
- Purpose: To analyze customer behavior related to churn and perform survival analysis predicting when and why customers leave the service.

| Customer ID | Gender | Payment Meathod | Tenure | Monthly Charges | Churn |
|---|---|---|---|---|---|
| 001 | Male | Electronic Check | 12 | 29.85 | Yes |
| 002 | Female | Mailed Check | 45 | 56.95 | No |
| 003 | Female | Bank Transfer | 2 | 53.85 | Yes |



As we can see the higher the tenure, the lesser the churn rate. This tells us that the customer becomes loyal with the tenure.



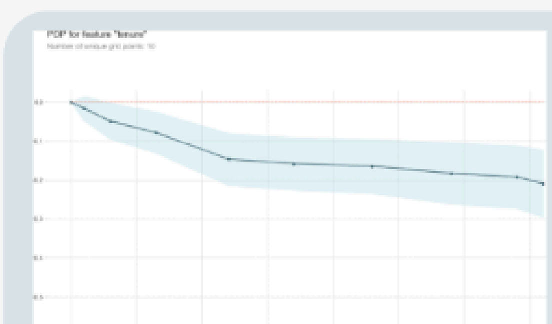

When the customers are new they do not opt for various services and their churning rate is very high. This can be seen in above plot.

## *Explainable AI modules using Random forest models-*

### Permutation Importance

| Weight | Feature |
|---|---|
| 0.0185 ± 0.0058 | InternetService_Fiber optic |
| 0.0064 ± 0.0088 | Contract_Two year |
| 0.0045 ± 0.0058 | OnlineSecurity |
| 0.0041 ± 0.0134 | Contract_One year |
| 0.0038 ± 0.0086 | PaymentMethod_Electronic check |
| 0.0037 ± 0.0071 | InternetService_No |
| 0.0028 ± 0.0094 | tenure |
| 0.0026 ± 0.0011 | OnlineBackup |
| 0.0020 ± 0.0078 | MonthlyCharges |
| 0.0010 ± 0.0014 | DeviceProtection |
| 0.0009 ± 0.0083 | PaperlessBilling |
| 0.0007 ± 0.0030 | TechSupport |
| 0.0004 ± 0.0032 | StreamingMovies |
| 0.0003 ± 0.0017 | gender |
| 0.0001 ± 0.0019 | PhoneService |

Permutation Importance shows feature importance by randomly shuffling feature values and measuring how much it degrades our performance.



### Partial Dependence plots

Partial dependence plot is used to see how churning probability changes across the range of particular feature. For example, in side graph of tenure group, the churn probability decreases at a higher rate if a person is in tenure group 2 compared to 1.

### Shap values

Shap values (Shapley Additive explanations) is a game theoretic approach to explain the output of any machine learning model. In above plot we can see that why a particular customer's churning probability is less than baseline value and which features are causing them.

**Core Equations:**

- **Random Forest**

$$\hat{p}(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M}\hat{p}_m(\mathbf{x})$$

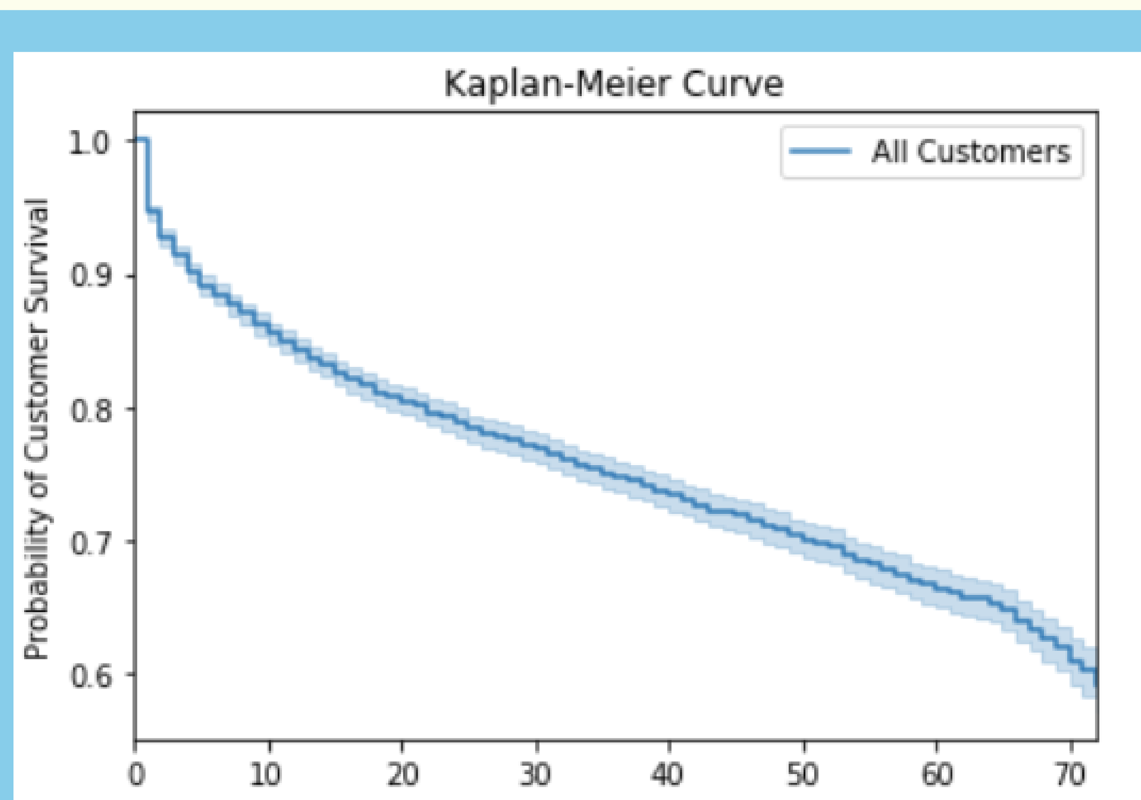
Splits chosen to minimize Gini impurity:

$$G = 1 - \sum_k p_k^2$$

- **Cox Proportional Hazards Model**

$$h(t|\mathbf{x}) = h_0(t)\exp(\beta^\top \mathbf{x})$$

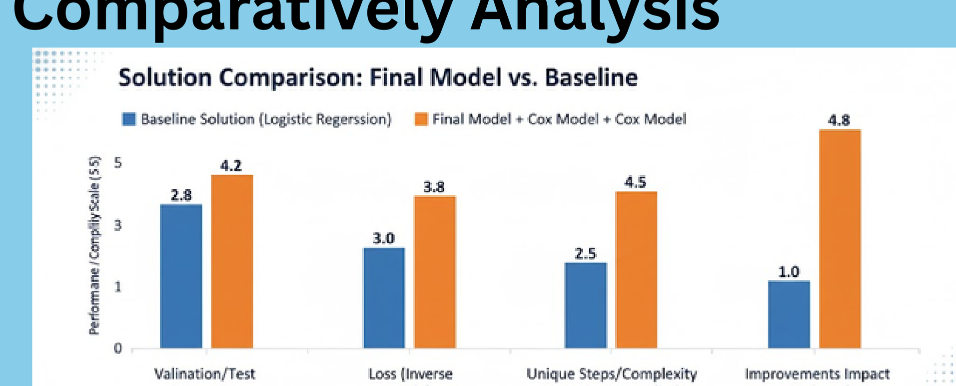
$$S(t|\mathbf{x}) = (S_0(t))^{\exp(\beta^\top \mathbf{x})}$$

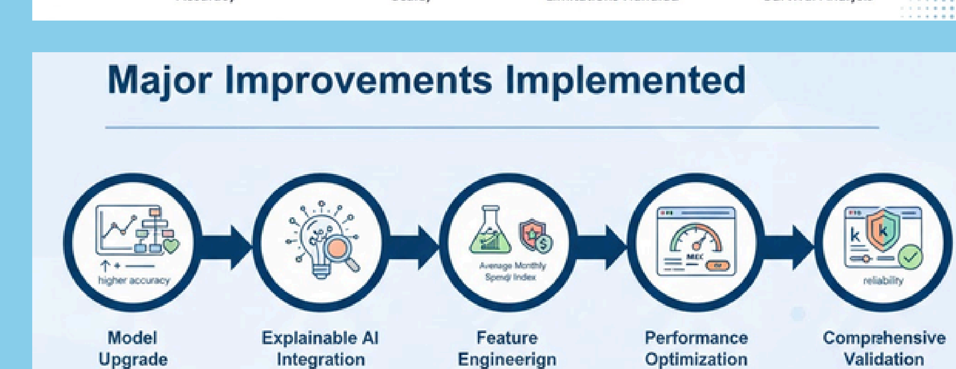



From above graph, we can say that

- AS expected, for telcom, churn is relatively low. The company was able to retain more than 60% of its customers even after 72 months.
- There is a constant decrease in survival probability probability between 3-60 months.
- After 60 months or 5 years, survival probability decreases with a higher rate.

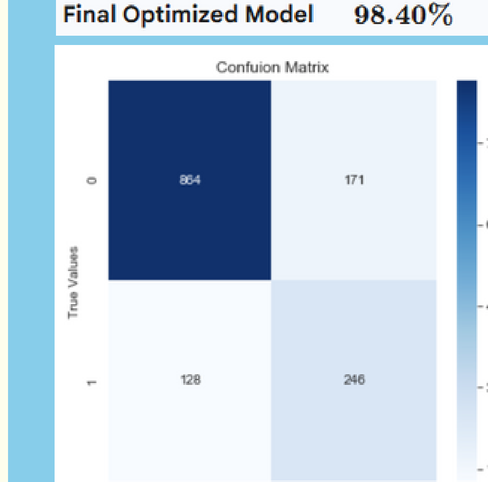
## Comparatively Analysis

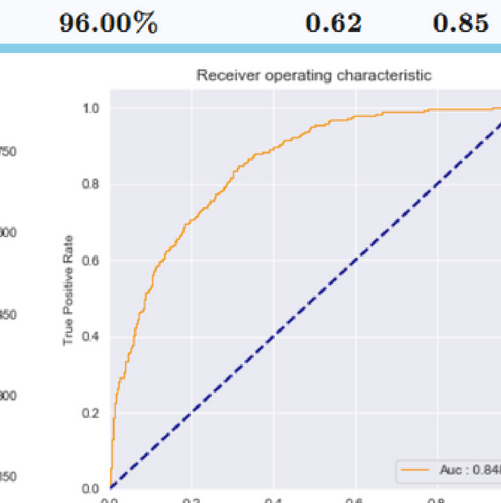



## Error Analysis

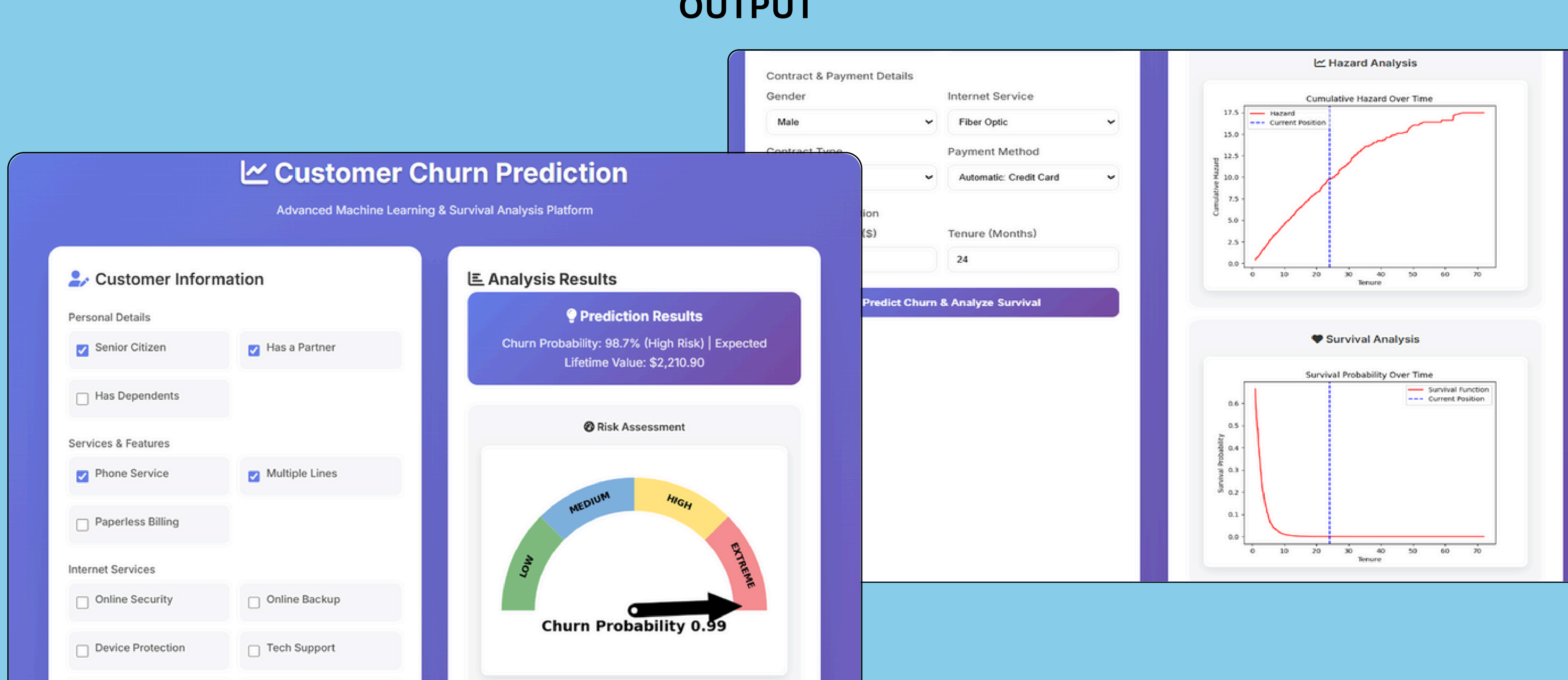
| Stage/Metric | Train Accuracy | Validation Accuracy | F1 Score | ROC-AUC |
|---|---|---|---|---|
| Initial Test Run | 92.50% | 88.00% | 0.55 | 0.80 |
| After Parameter Tuning | 97.00% | 95.50% | 0.60 | 0.84 |
| **Final Optimized Model** | **98.40%** | **96.00%** | **0.62** | **0.85** |




The final model resulted in 0.62 F1 score and 0.85 ROC-AUC.

## OUTPUT




## Results Evaluation

Performance Metrics

| Metric | Result | Interpretation |
|---|---|---|
| Accuracy | 0.94 | The model correctly classified most of the test samples on the dataset. |
| Precision | 0.94 | Almost all customers predicted to churn were actual churners. |
| Recall | 0.98 | The model identified all churners correctly. |
| Mean Squared Error (MSE) | 0.06 | Little to no error between predicted and actual values. |

**Exceptional Detection:** The model correctly identifies almost all potential churners. (98% recall)
**Minimizes Risk:** This minimizes costly False Negatives
**Highly Accurate:** The predictions are reliable and trustworthy.(94% Accur)
**Effective Intervention:** The results allow the business to target retention efforts precisely.