PURWADHIKA DIGITAL TECHNOLOGY SCHOOL

Machine Learning & Cloud Computing **Capstone Project**

TELCO CUSTOMER CHURN PREDICTION

Randy Prawiro
DTIDS-0106

Agenda

Business Problem Formulation

Churn causes short-term revenue to plummet.

Churn

- Churn is when a customer chooses to stop using a company's products or services.
- In effect, it's when a customer ceases to be a customer.

Why We Should Be Concerned

- In the telecom industry, churn causes disruptions to short-term cash flow.
- Results in the slowing down or even a complete halt of operations
- Places the company in a difficult liquidity situation.

Our Company Had a Churn Rate of 26.54%!

Back to Agenda

Machine Learning Approach

How can we utilize our data to build the model?

How can the model be used to solve our business problem?

Goal # 1

Understand our dataset and build a machine learning model for classification that aims to predict customer churn before they happen.

Goal # 2

The model labels the customers as either churning or not churning based on their subscription patterns, transaction history, and personal characteristics.

Goal # 3

Train our model on historical churn data from our dataset, and increase its performance by optimizing its test data predictions.

Goal # 4

Deliver personalized retention strategies for customers who are predicted to churn, such as discount offerings and recommendations for alternative services.

Back to Agenda

EDA & Data Preprocessing

Start by understanding the characteristics of our dataset and transform it into an ideal format in order for our model to learn effectively.

Exploratory Data Analysis & Data Preprocessing	
8 categorical features, 2 numerical features	No extreme class imbalances in the categorical columns so resampling will not be conducted
The fiber optic internet service, month-to-month contract subscription, and paperless billing features were associated with significant amounts of customer churn	No global, contextual, nor group outliers in the numerical features
Customers who churned were found to have shorter tenure and higher monthly charges compared to those who didn’t	One hot encoder used to convert all categorical features into numerical features
77 duplicate entries were dropped to prevent overfitting	Feature selection using SelectKBest with the chi squared scoring function and k = 6
No entries with missing values	Standard scaler used to scale numerical features to have a mean of 0 and a standard deviation of 1

Back to Agenda

Model Benchmarking

Initial model testing to check the performance of the three best performing models

Model Name	Testing Score	Training Score	Difference
Logistic Regression	0.556	0.538	0.018235
Gradient Boosting	0.500	0.539	-0.039874
Extreme Gradient Boosting	0.484	0.521	-0.037398

logreg, knn, dtree, rf, ada, gbc, xgb were used to compare test set performance with each other.

Seven benchmark models were used on the basis of the f2 score. The three best performing models ranked by order of the highest evaluation metric are Logistic Regression, Gradient Boosting, and finally Extreme Gradient Boosting.

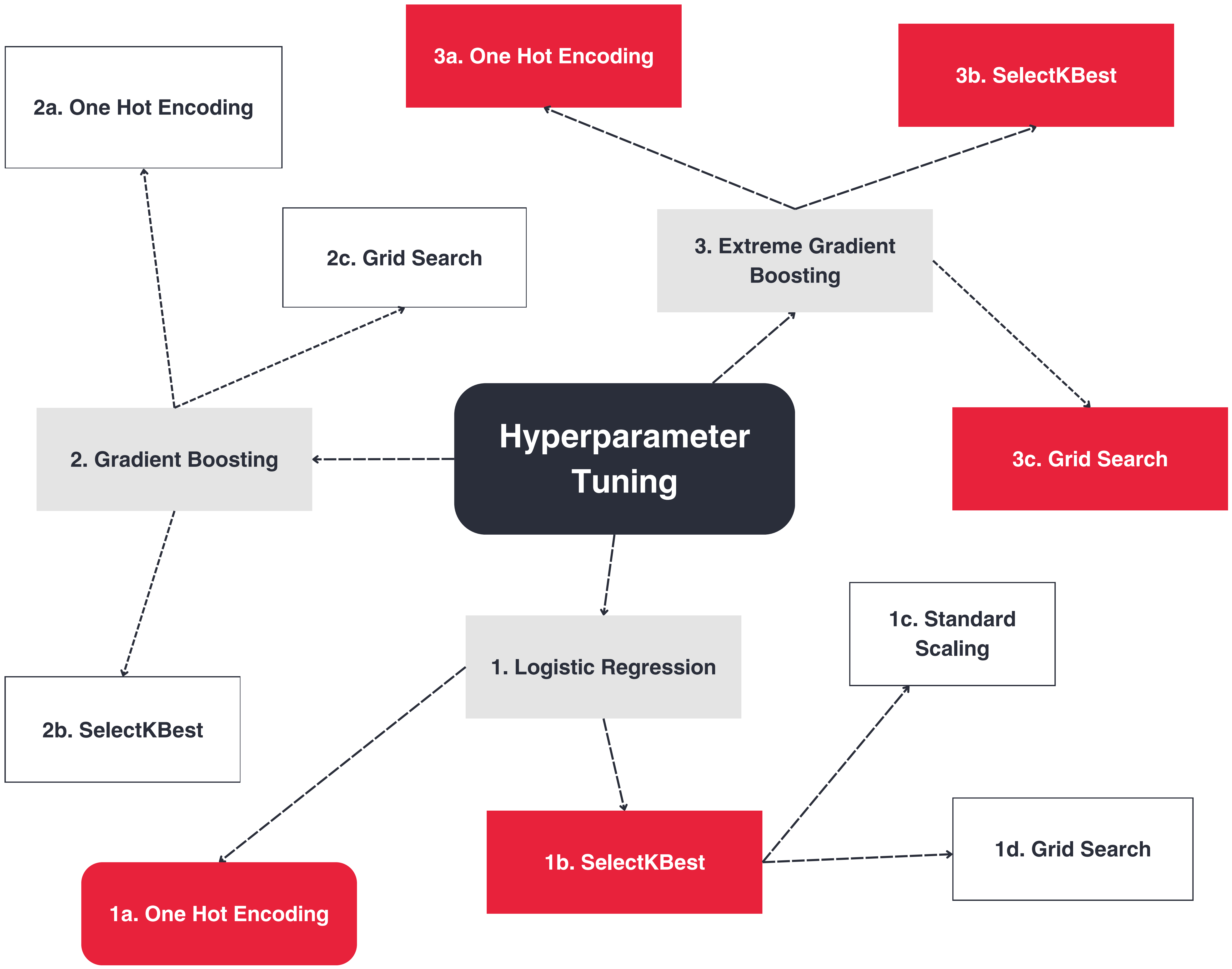
Back to Agenda

Hyperparameter Tuning

Logistic Regression:
penalty, C, solver, max_iter, multi_class, tol, class_weight

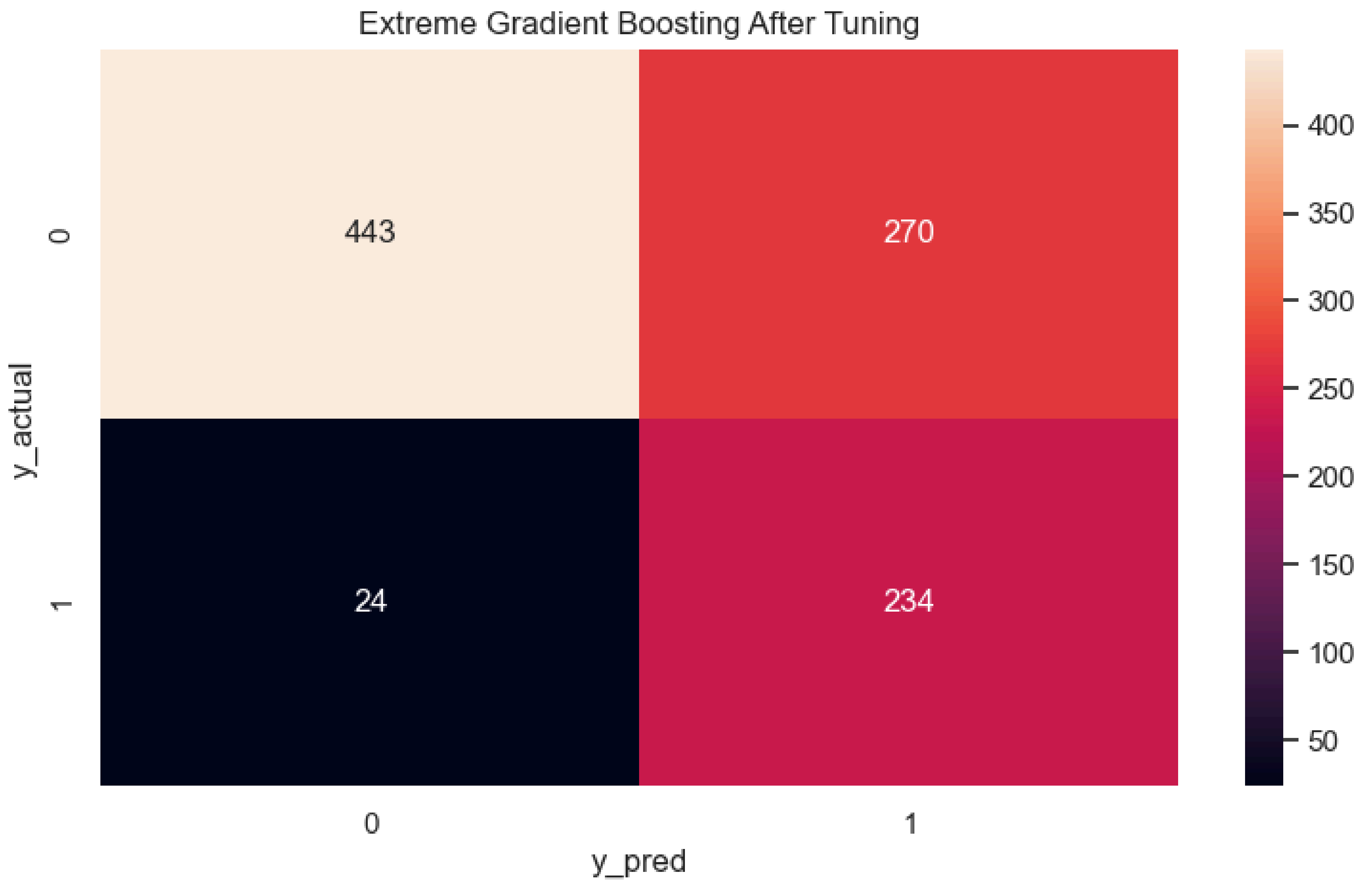
Gradient Boosting:
learning_rate, n_estimators, max_depth, min_samples_split, min_samples_leaf, max_features, subsample

Extreme Gradient Boosting:
n_estimators, max_depth, learning_rate, gamma, reg_alpha, reg_lambda, scale_pos_weight



Final Model & Evaluation:
Tuned XGBoost

Best tuned model based on highest recall & f2 score



Model Performance:

```
Extreme Gradient Boosting After Tuning
Recall_Score :  0.9069767441860465
F2_Score     :  0.76171875
```

Model Advantage:

- Very high recall for the positive/churn class
- Very good at identifying churn cases when they actually occur

Model Disadvantage:

- Poor precision for the positive/churn class
- Bad at making churn predictions that turned out to actually be churn cases

Back to Agenda

Final Model & Evaluation:
Tuned XGBoost

- **gamma:** Controls the regularization of tree splits.
- **learning_rate (eta):** Determines the step size during gradient boosting.
- **max_depth:** Specifies the maximum depth of the decision trees.
- **n_estimators:** Number of boosting rounds or trees.
- **reg_alpha:** L1 regularization term (Lasso regularization).
- **reg_lambda:** L2 regularization term (Ridge regularization).
- **scale_pos_weight:** Adjusts the balance between positive and negative weights in imbalanced datasets.

	precision	recall	f1-score	support
0	0.95	0.62	0.75	713
1	0.46	0.91	0.61	258
accuracy			0.70	971
macro avg	0.71	0.76	0.68	971
weighted avg	0.82	0.70	0.71	971

Tuned XGBoost Best_Params :
{'gamma': 0.4, 'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 100, 'reg_alpha': 20, 'reg_lambda': 10, 'scale_pos_weight': 5}

Cost-Benefit Analysis

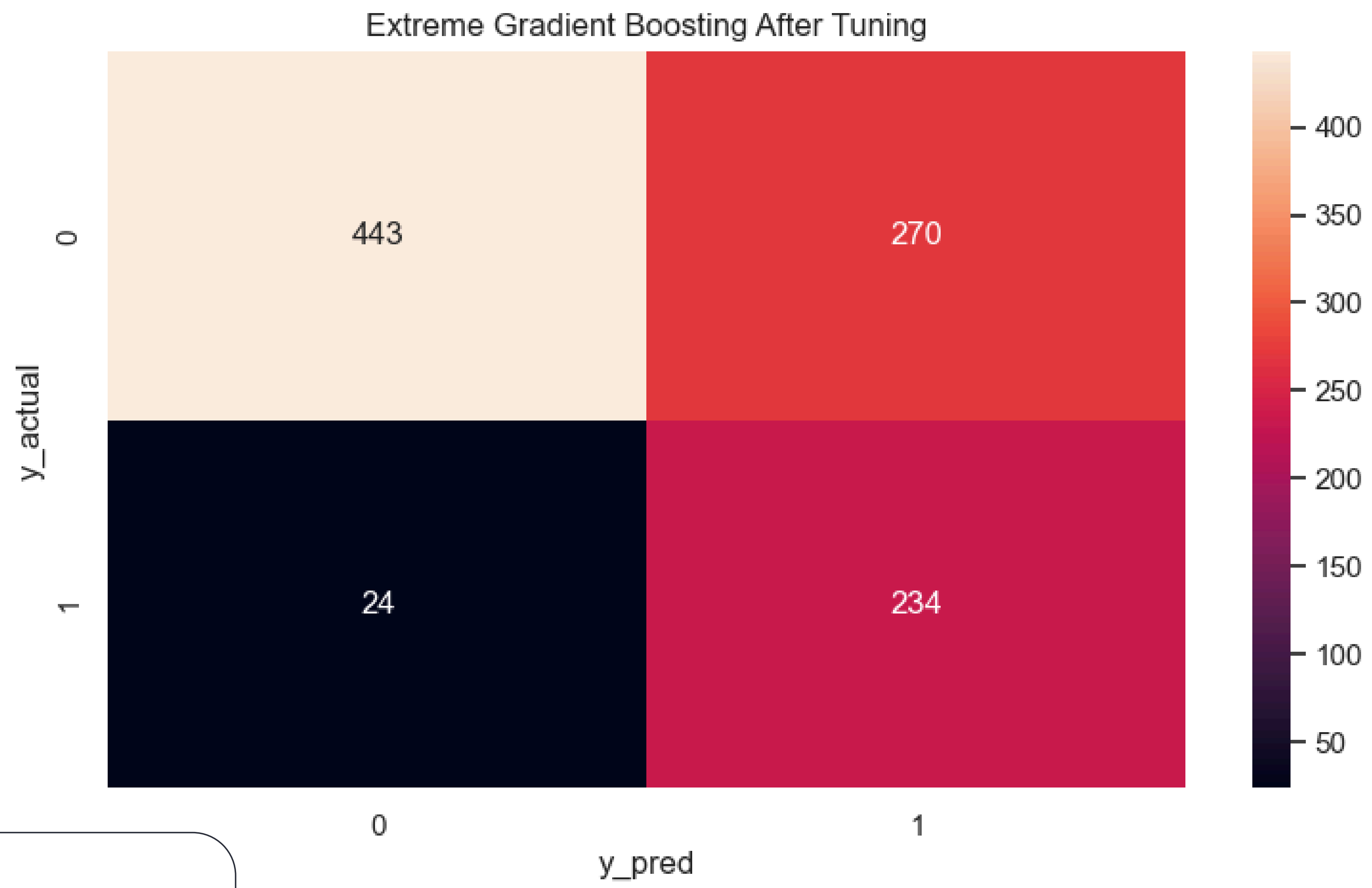
Comparing model implementation vs no implementation

Net Revenue Saved:

- The total potential revenue that could be saved from churn is $96,532
- The total cost of false predictions is the cost of False Positive + cost of False Negative
- Equals 13.10 * FP + (1 - recall) * 96,532
- This results in a total loss of $12,185 due to False predictions
- The net saved revenue is $84,347

Impact of Model Implementation:

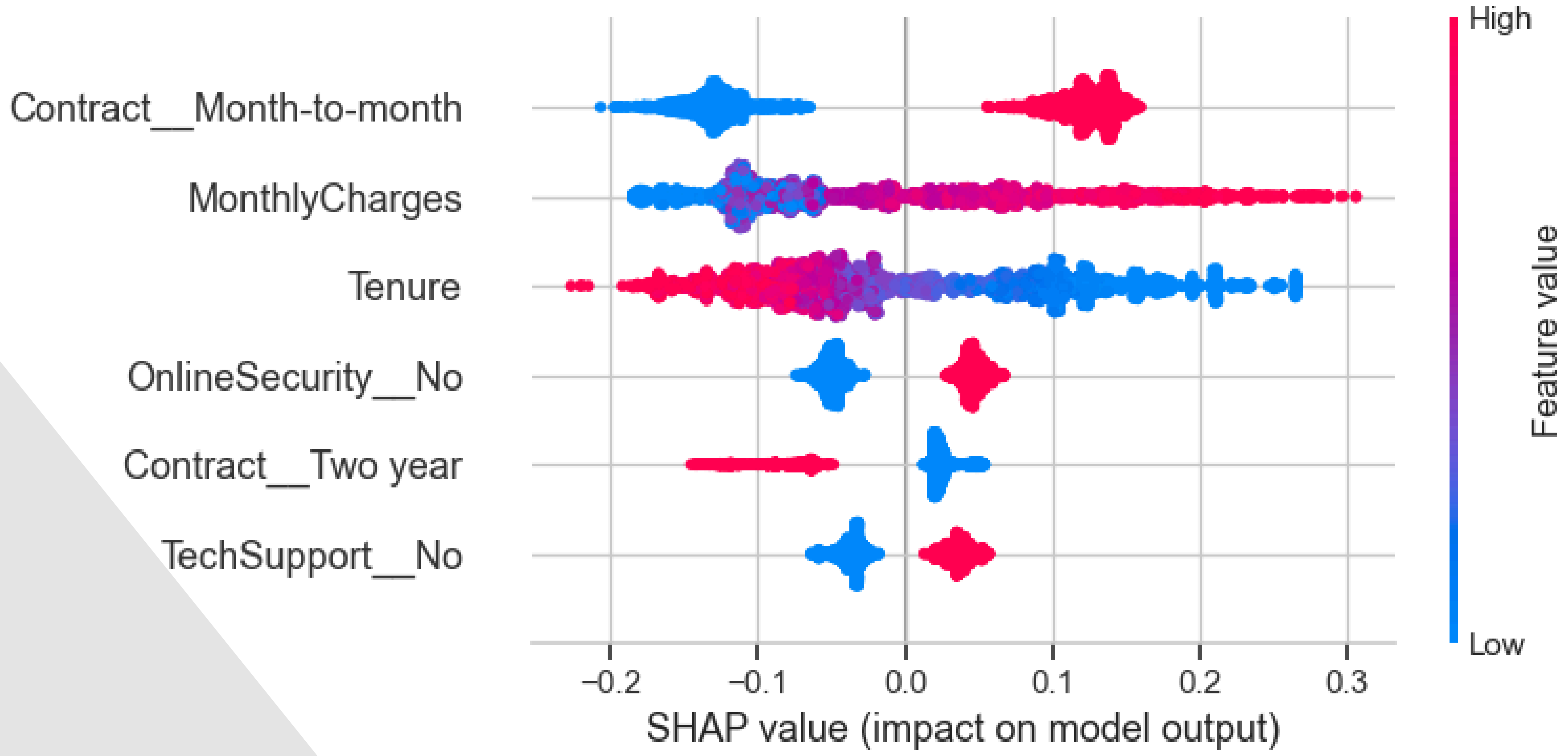
- Reduce churn costs by 87.38% ($84,347/$96,532)
- Reduce the churn rate from 26.54% to only 2.47% (24/(24+234+443+270))



****The equations used in the cost-benefit analysis were derived during the EDA and recommendations sections in the jupyter script and are more comprehensively discussed there**

Back to Agenda

Explainable AI



Six best features were selected for the final model.

---

4 binary feature importances in the model:

Month-to-month contract, no online security, two year contract, and no tech support

---

2 other numerical feature importances in the model:

Monthly charges and tenure

Back to Agenda

Cloud Computing Demonstration

GCP Environment Setup & Cloud Prediction

Conclusion & Recommendations

- The final model was able to identify 91% of all churn cases but would misclassify 38% of non-churning customers.
- It was observed that by implementing this model, the telecom company was able to reduce its churn rate from 26.54% to only 2.47% and reduce the cost resulted by churn by 87.38%..
- It is recommended to deliver personalized retention strategies to customers who are predicted to churn and invest in the company's services to prevent churn.
- Cloud deployment would make our model more accessible to a wider range of users.

Back to Agenda