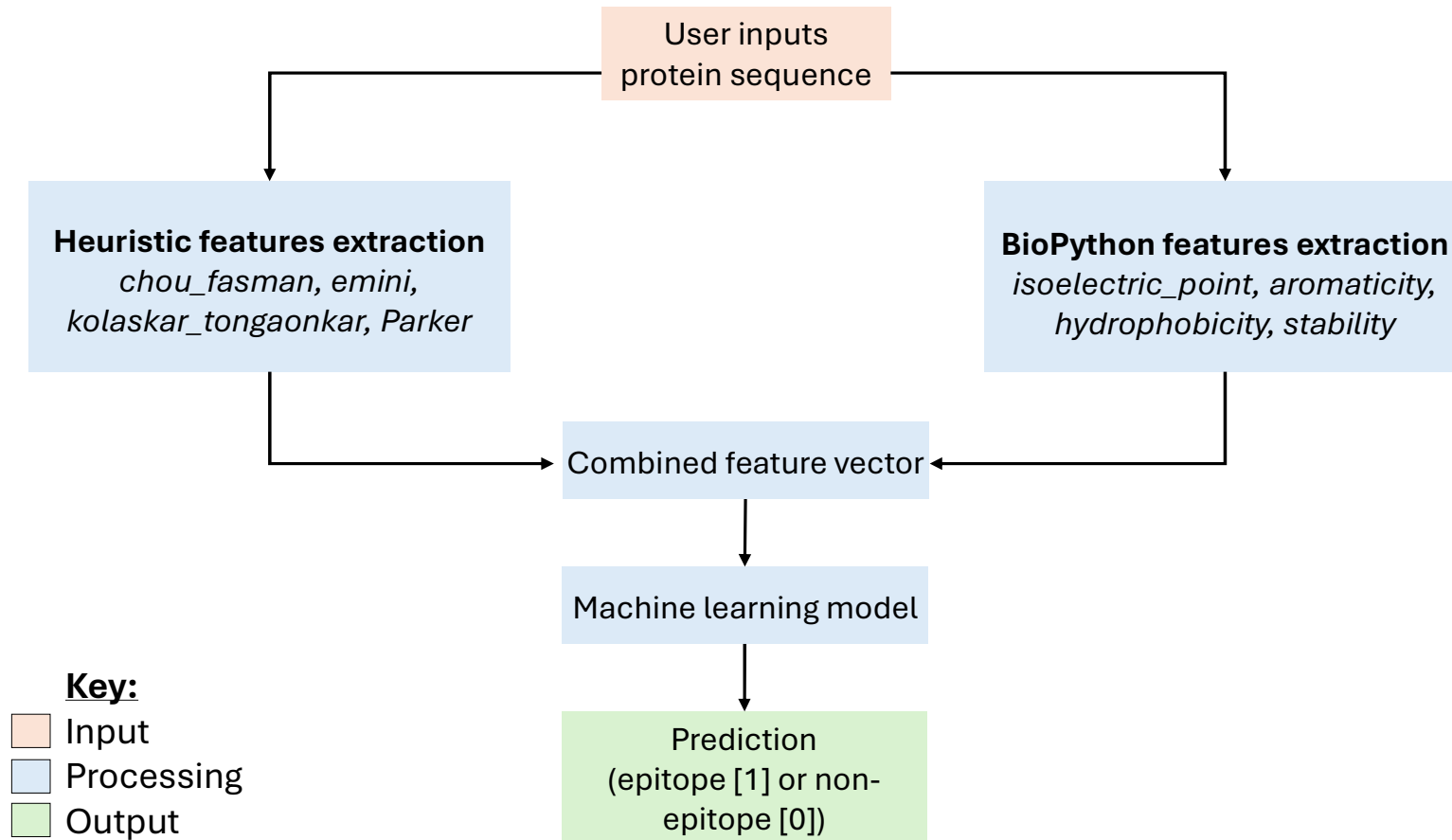


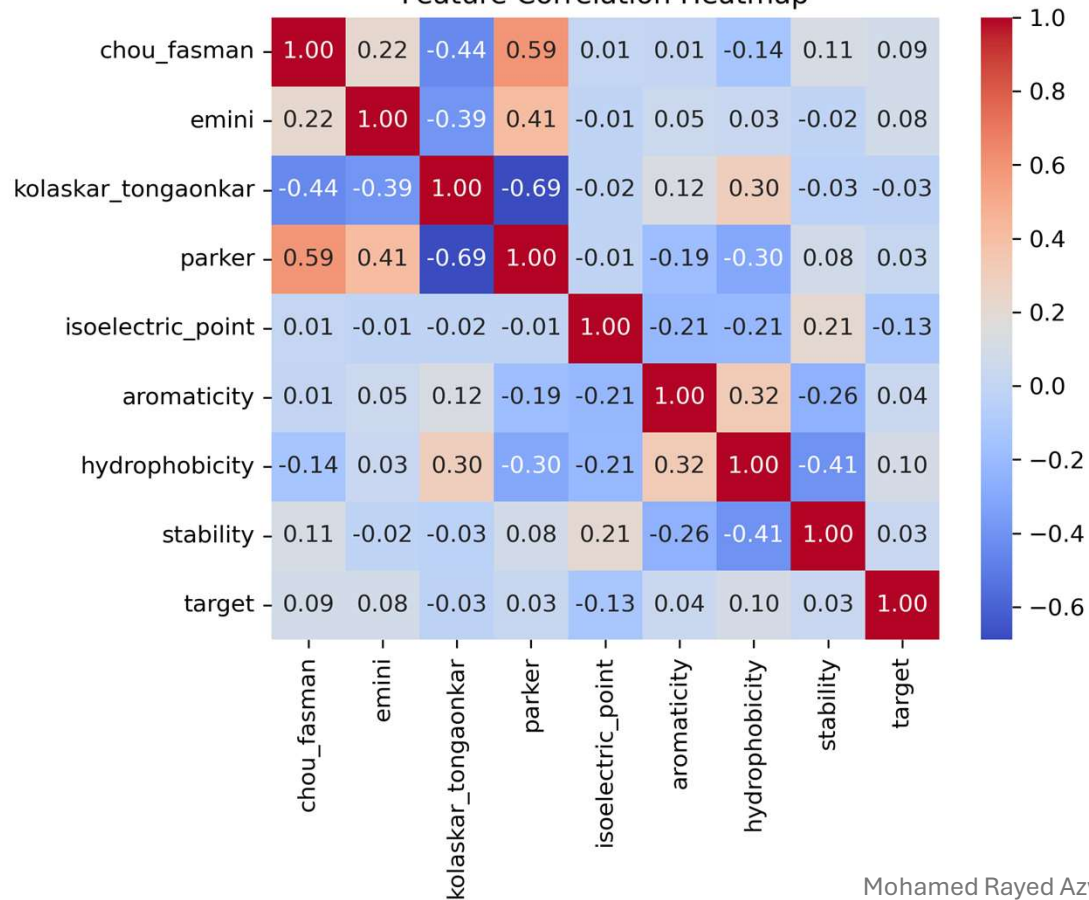
# Machine Learning Pipeline

*B-Cell epitope prediction for COVID-19 / SARS*



# Correlation Heatmap

Feature Correlation Heatmap

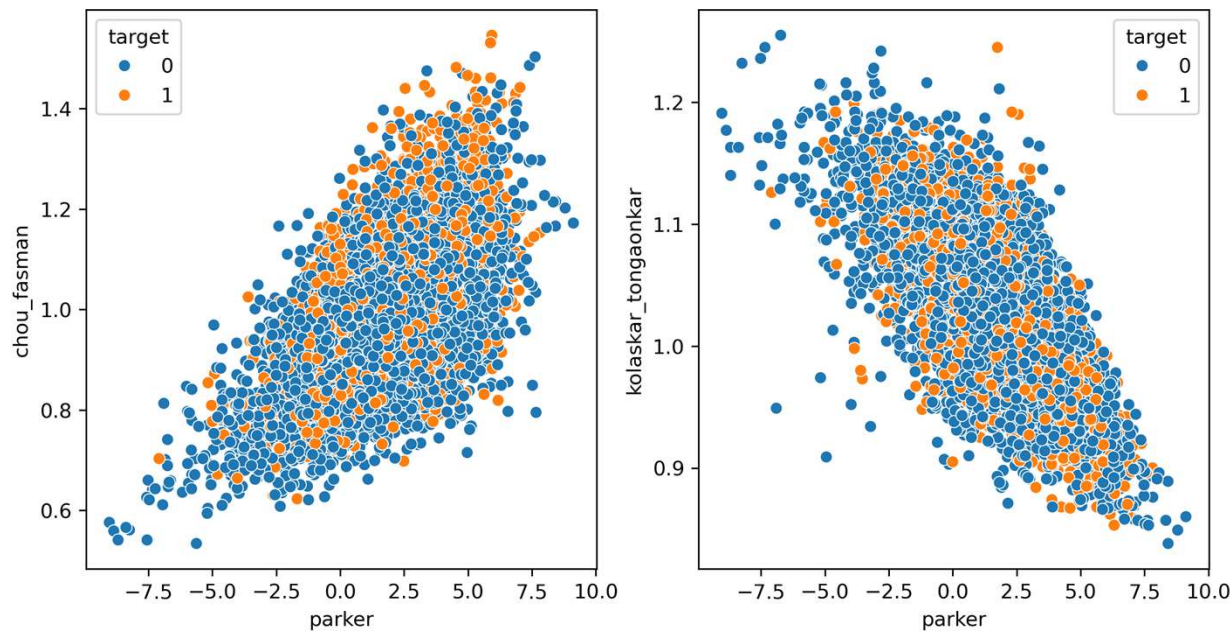


Parker hydrophilicity and Chou-Fasman scores show a moderate **positive correlation (+0.59)**, suggesting regions that are hydrophilic tend to also favour beta-turn structures

Parker hydrophilicity and Kolaskar-Tongaonkar scores show a **strong negative correlation (-0.69)**, suggesting that highly hydrophilic regions are less likely to be predicted as antigenic by Kolaskar-Tongaonkar.

# Comparative Feature Relationship

*Parker vs Chou-Fasman and Kolaskar-Tongaonkar*

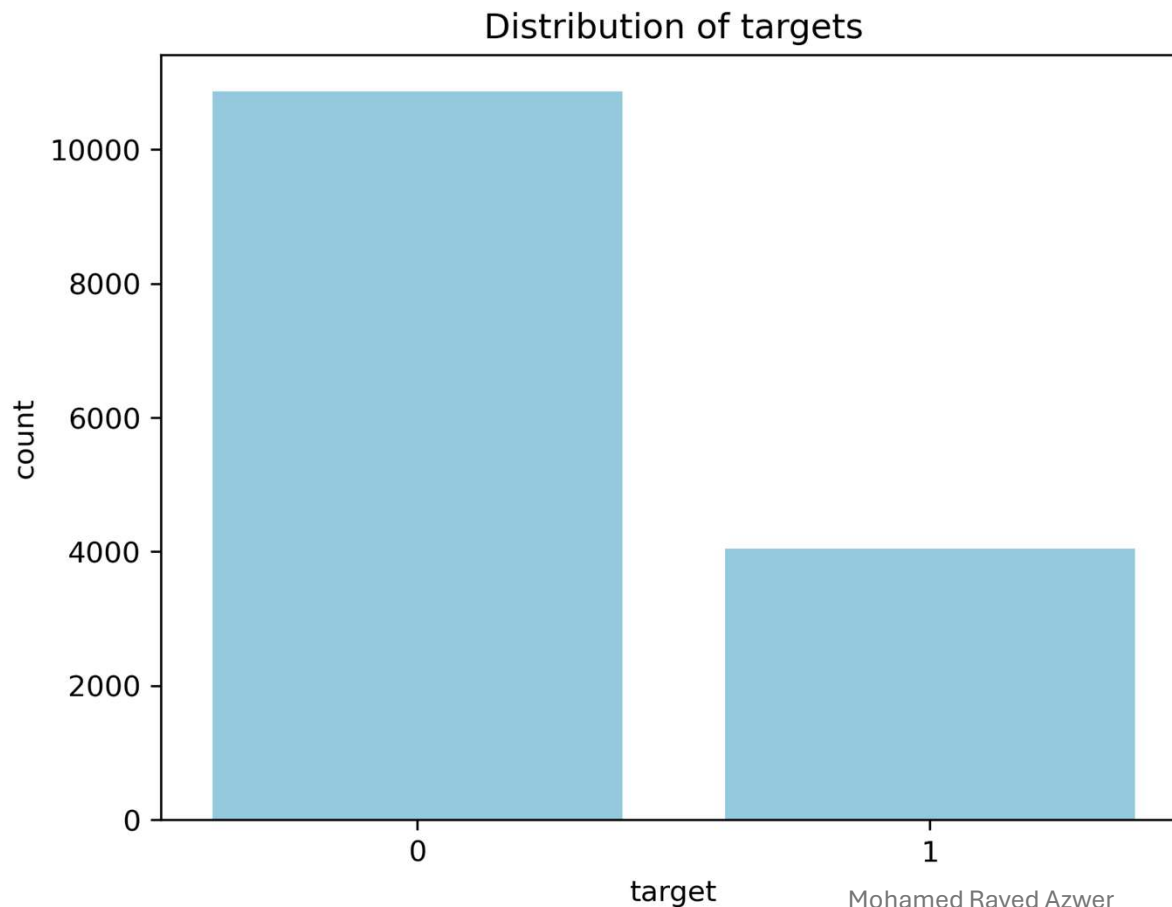


The scatter plots show how **hydrophilicity (Parker)** interacts with two other feature scores.

**Parker and Chou-Fasman** showed a positive trend, meaning hydrophilic residues often coincide with regions having a higher Chou-Fasman beta-turn propensity, which represents the probability of a peptide forming flexible, surface-exposed structures.

**Parker and Kolaskar-Tongaonkar** displayed a negative correlation, suggesting these two heuristics capture distinct physicochemical aspects of epitope behaviour.

# Training Dataset Class Distribution For Epitope Prediction



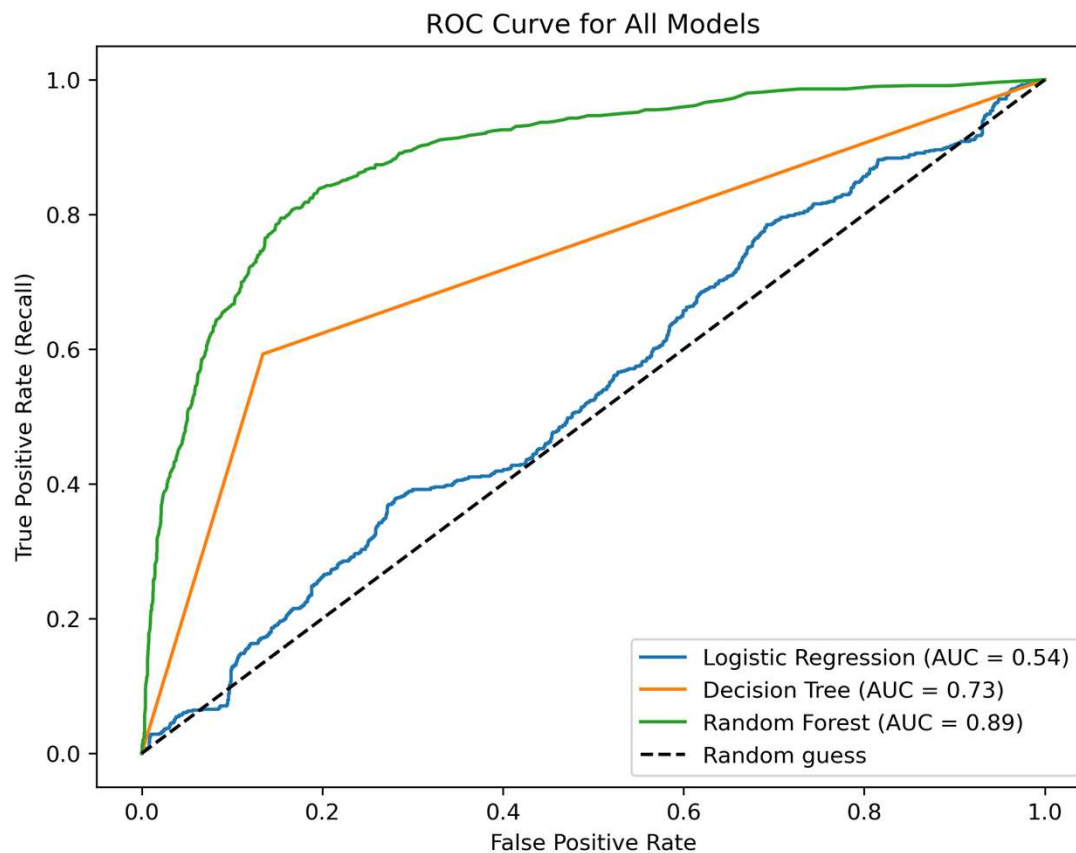
Mohamed Rayed Azwer

The dataset used in this machine learning model for training was **unbalanced** between targets, 0 (non-epitope), and 1 (epitope)

To ensure a fair evaluation, **stratified train-test splitting was employed**

```
target
0      10865
1       4042
Name: count, dtype: int64
```

# ROC Curve Analysis of Classifiers for Epitope Prediction



Three classifiers were evaluated in the machine learning pipeline: **Logistic Regression, Decision Tree, and Random Forest.**

Amongst them, the **Random Forest model performed best**, achieving an Area Under the ROC Curve (AUC) of 0.89, indicating strong predictive ability.

In contrast, **Logistic Regression showed the weakest performance** with an AUC of 0.54, which is roughly equivalent to random guessing.