



## Introduction

- Online dating is growing, yet many users **struggle to find meaningful matches**.
- Existing algorithms **lack depth—they miss deeper compatibility factors**.
- This project combines **psychology** and **ML** to **improve dating match predictions**.

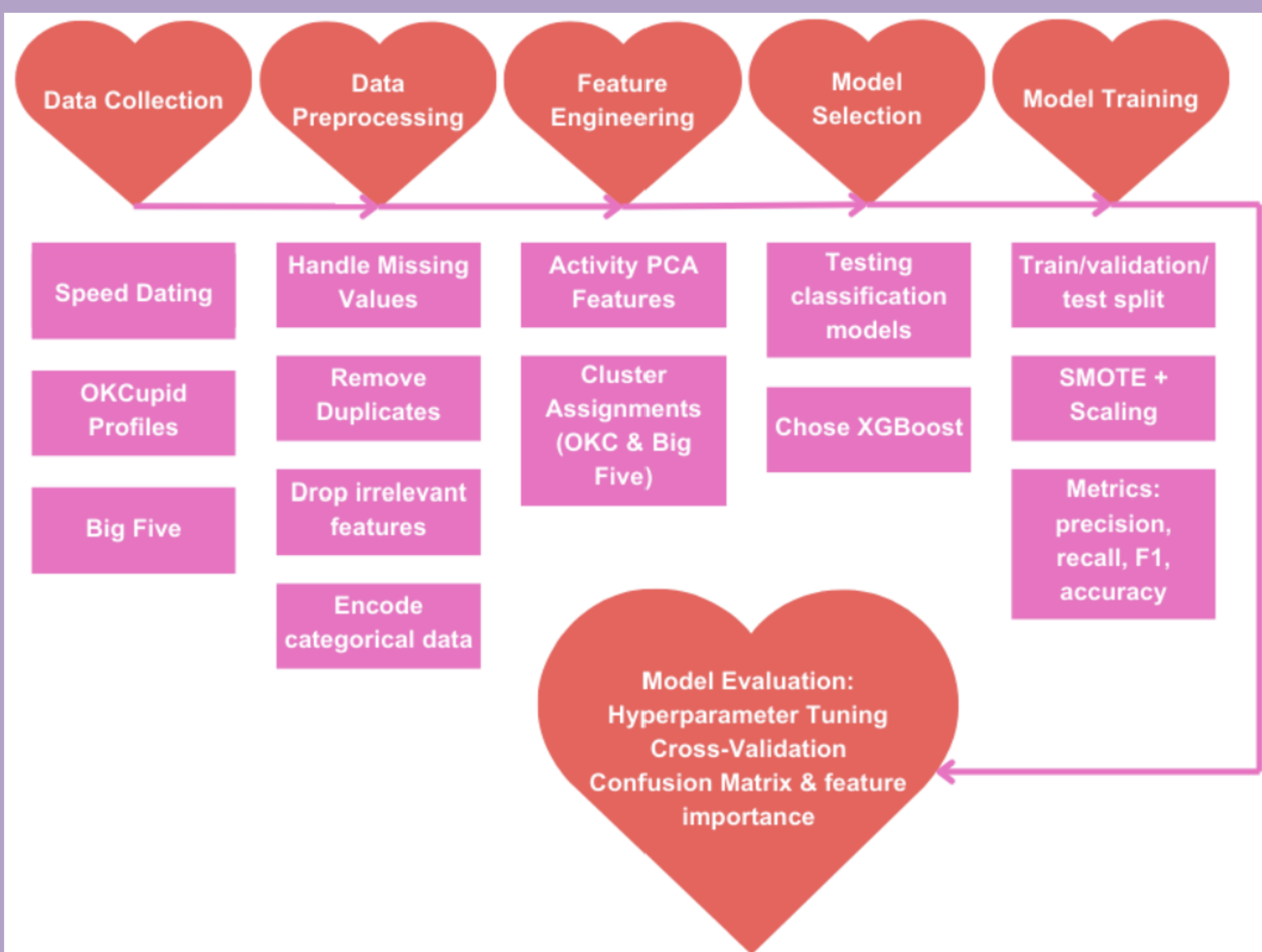


## Approach

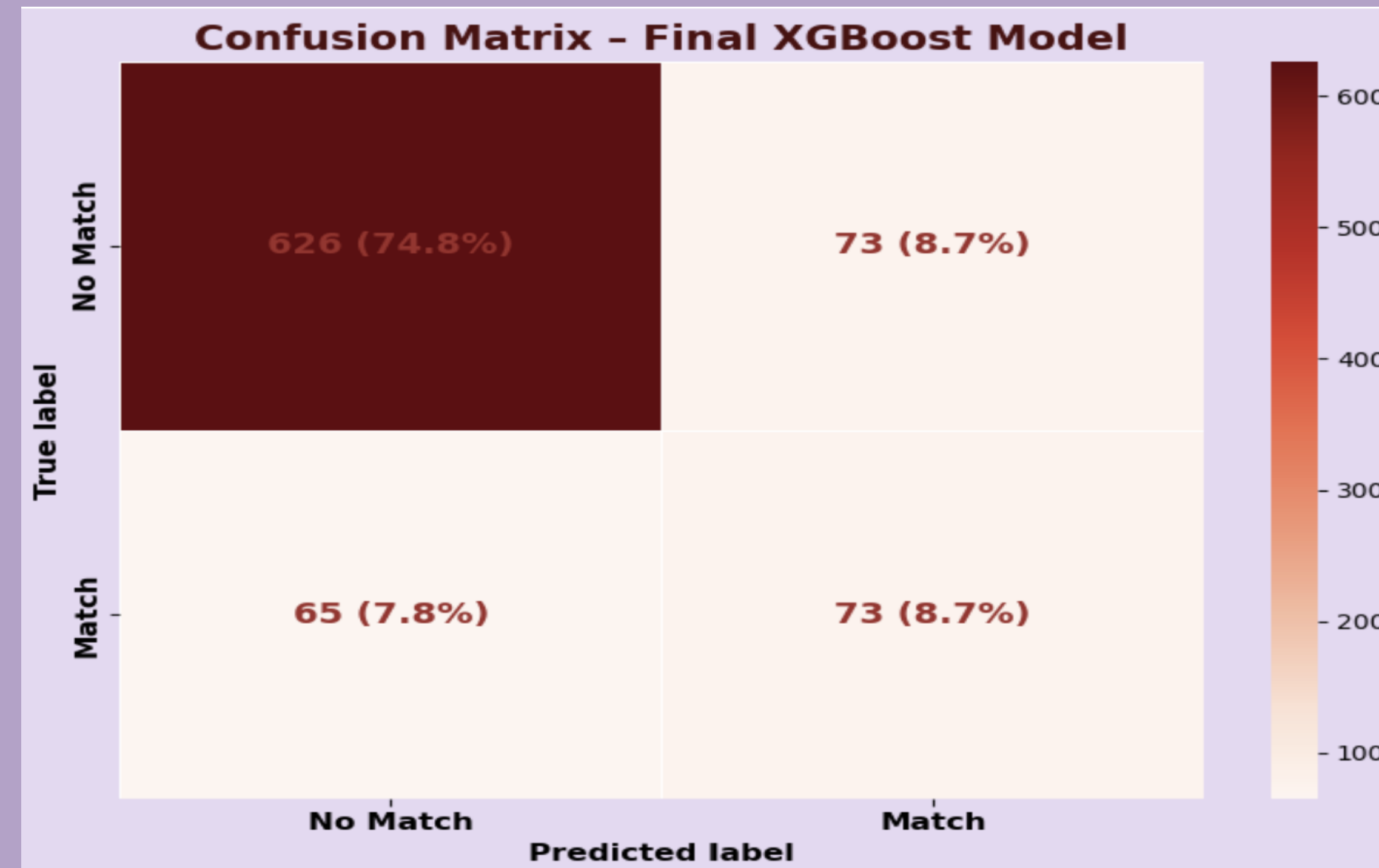
- Collected and merged data from three sources: **Speed Dating**, **OKCupid**, and **Big Five personality** datasets
- Created **clusters** from **personality profiles** and **dating preferences**
- Engineered new features using **PCA** to represent user **lifestyle** and **activity patterns**
- Balanced class distribution using **SMOTE** and scaled features with **StandardScaler**
- Trained an **XGBoost** model to **predict** mutual matches based on **behavioral**, **demographic**, and **psychological data**



## Workflow

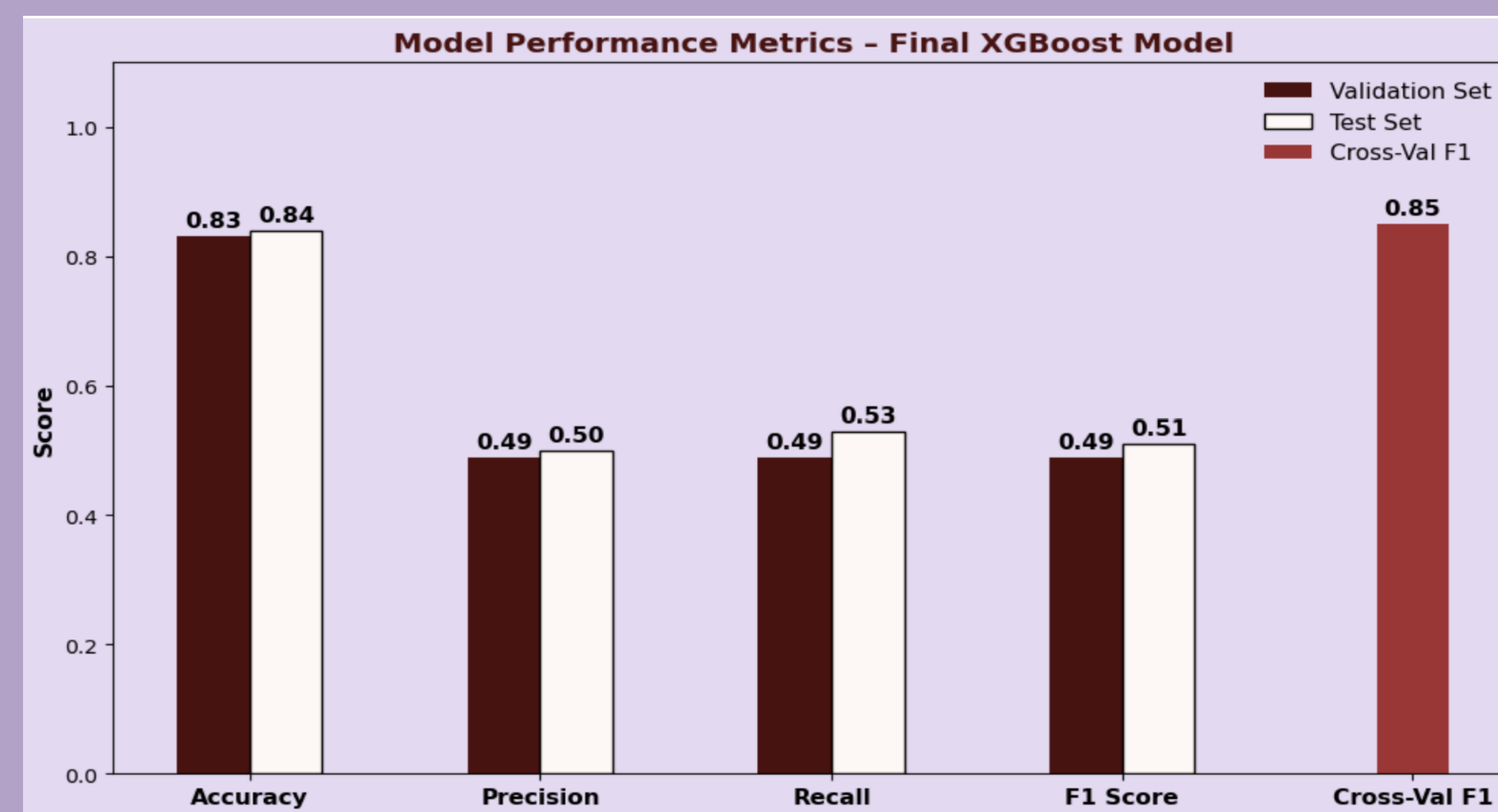


## Results



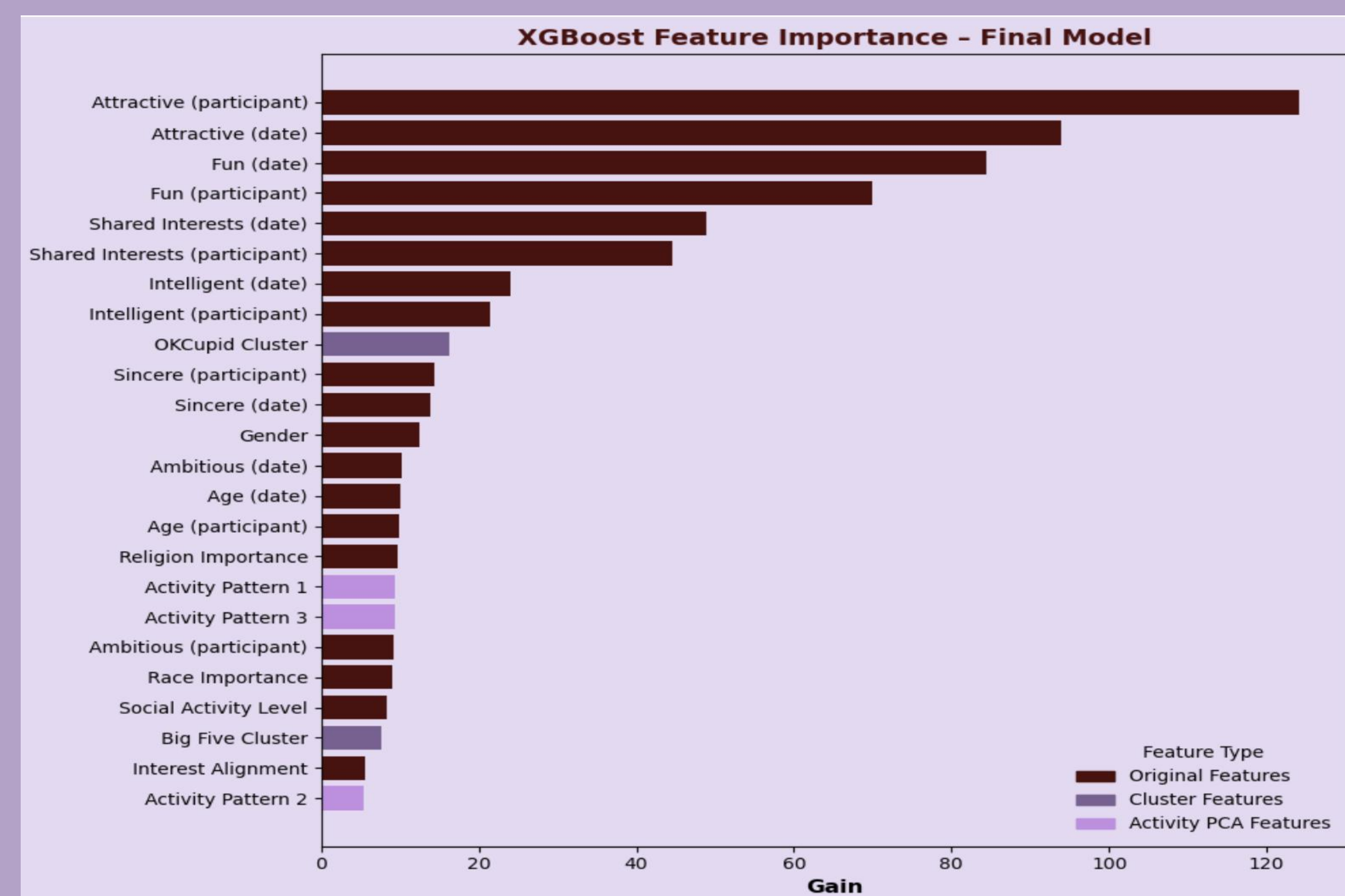
### Confusion Matrix:

- The model correctly **identifies non-matches most often (74.8%)**, with **fewer true matches (8.7%)** and a **small portion of errors (16.5% total misclassified)**.



### Model Metrics:

- Model shows **strong and consistent performance** across validation and test sets. **High cross-validated F1 score (0.85)** highlights reliable **generalization** and **robustness**.



### Feature Importance:

- Attractiveness** and **fun** ratings were the **strongest predictors** of matches, but **personality clusters** and **activity-based PCA features** also added valuable predictive power, highlighting the **importance** of both **preferences** and **behavior**.

## Conclusion

- Compatibility is complex:** It's best predicted using a mix of preferences, personality traits, and shared interests.
- Psychological alignment helps:** Personality-based clusters added strong predictive value.
- Attractiveness isn't everything:** Fun, intelligence, and interests also mattered in predicting matches.
- Addressing imbalance matters:** SMOTE and class weighting improved recall for real matches.
- Model is reliable:** Similar performance across validation and test sets confirms strong generalization.
- Smarter dating is possible:** ML + psychology can improve dating platforms and deepen our understanding of relationships.

## Limitations and Lessons Learned

- Small sample** of mutual matches **limited model balance**.
- Subjective rating** data introduced **noise**.
- Merging datasets** from different sources required **heavy cleaning** and **assumptions**.

## Potential Future Work

- Include more behavioral features** (e.g., messaging content, eye contact, vocal tone) to capture deeper aspects of compatibility
- Explore neural network models** to evaluate non-linear and interaction-heavy relationships
- Deploy an API prototype** for match scoring based on user profiles and psychology clusters
- Conduct real-time testing** with dating app users to evaluate recommendation performance

## References

- Speed Dating Experiment Dataset. Kaggle.*
- OKCupid Profiles Dataset. Kaggle.*
- Big Five Personality Traits Dataset. Kaggle.*

**GitHub Repository:**

