Online Signature Verification Using Machine Learning

- Basel Al-Raoush
- Supervisor: Dr. Mohammed Saleem
- Budapest University of Technology and Economics

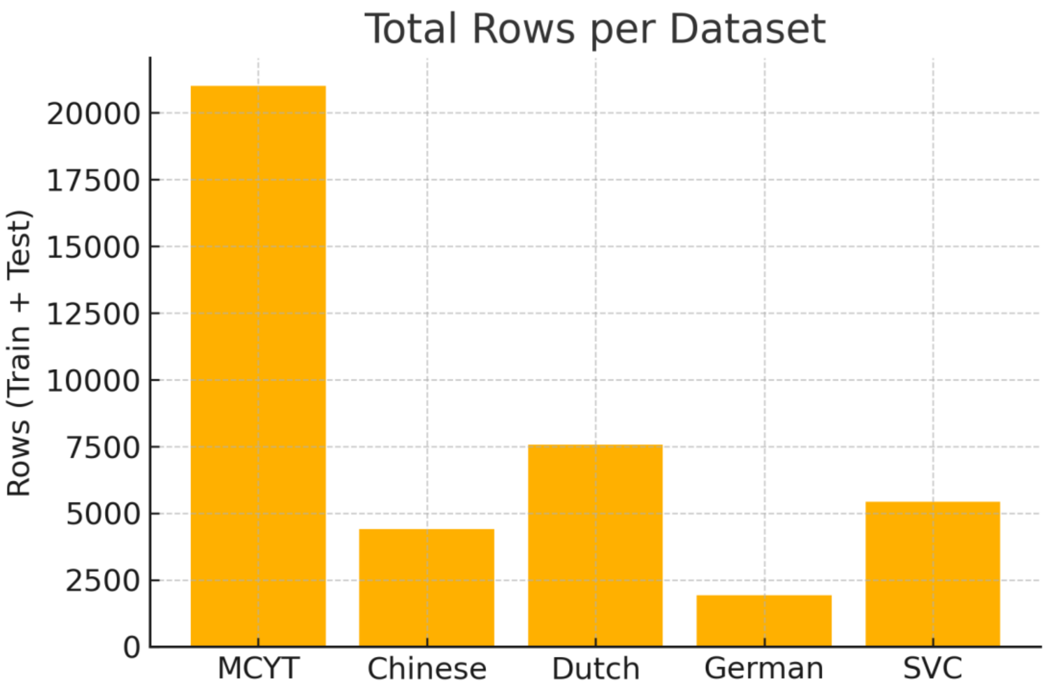
MŰEGYETEM 1782

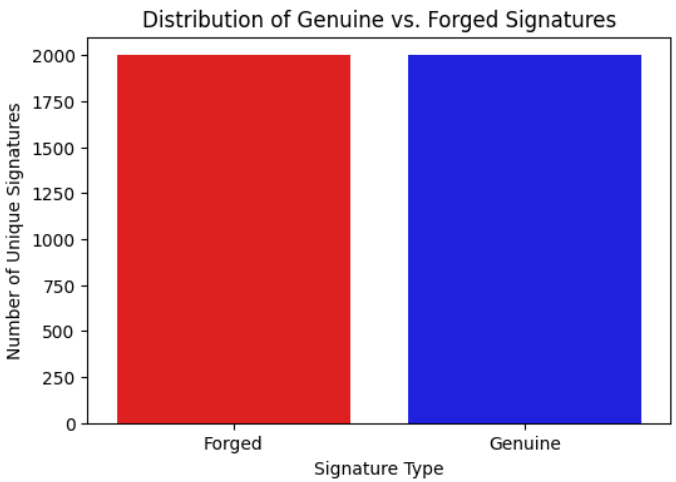
What is Online Signature Verification?

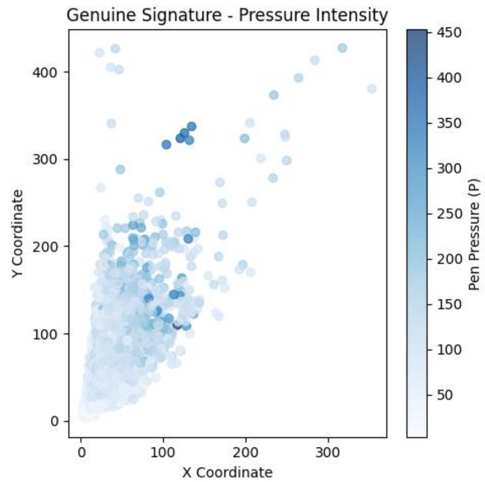
- Biometric method for identity verification
- Dynamic traits: pressure, tilt, direction
- Harder to forge than static (offline) signatures
- Used in banking, digital signing, legal access

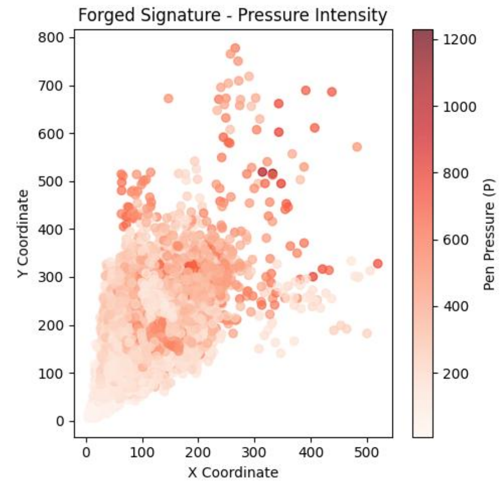
Datasets Used

- Main dataset: MCYT (100 users, real + forged)
- Additional: Chinese, Dutch, German, SVC datasets
- All datasets pre-split (train/test)







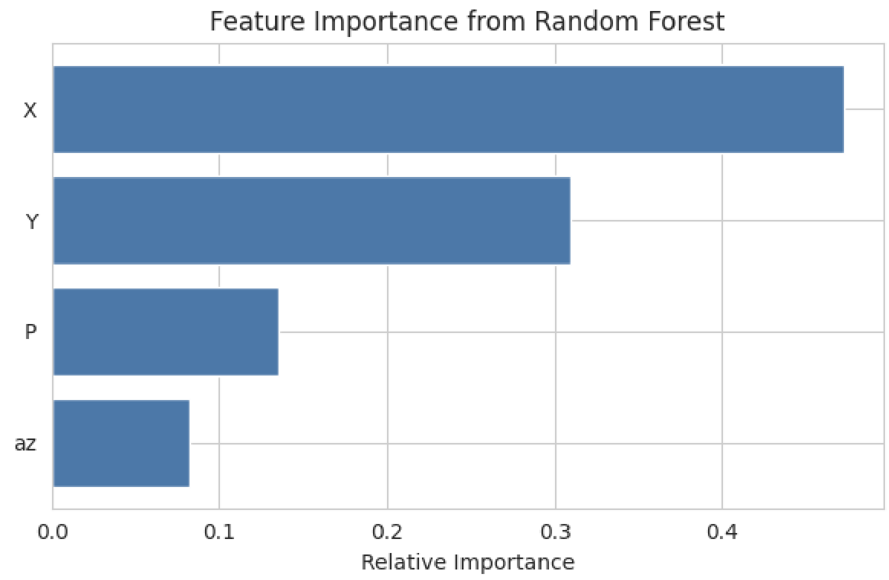


Insights from EDA

- Forged signatures = higher pressure, more variation
- Balanced class distribution, clean data
- Azimuth: more scattered in forgeries

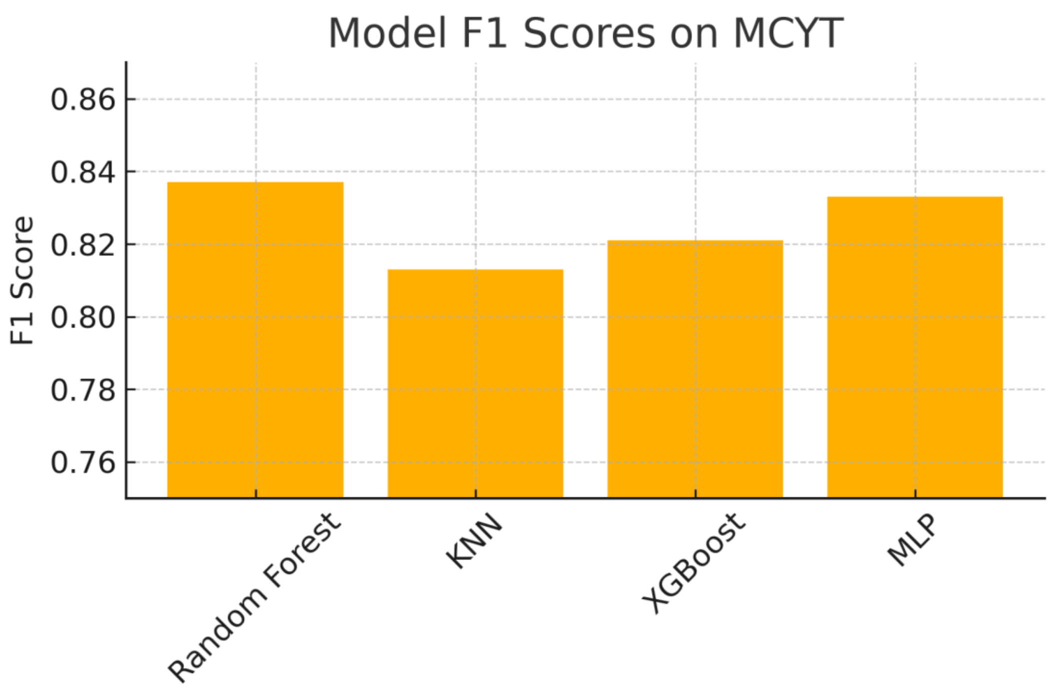
Features & Preprocessing

- Key features: pressure (P), azimuth (az), X/Y motion
- Extracted: mean, std, min, max
- Standardized features (z-score)



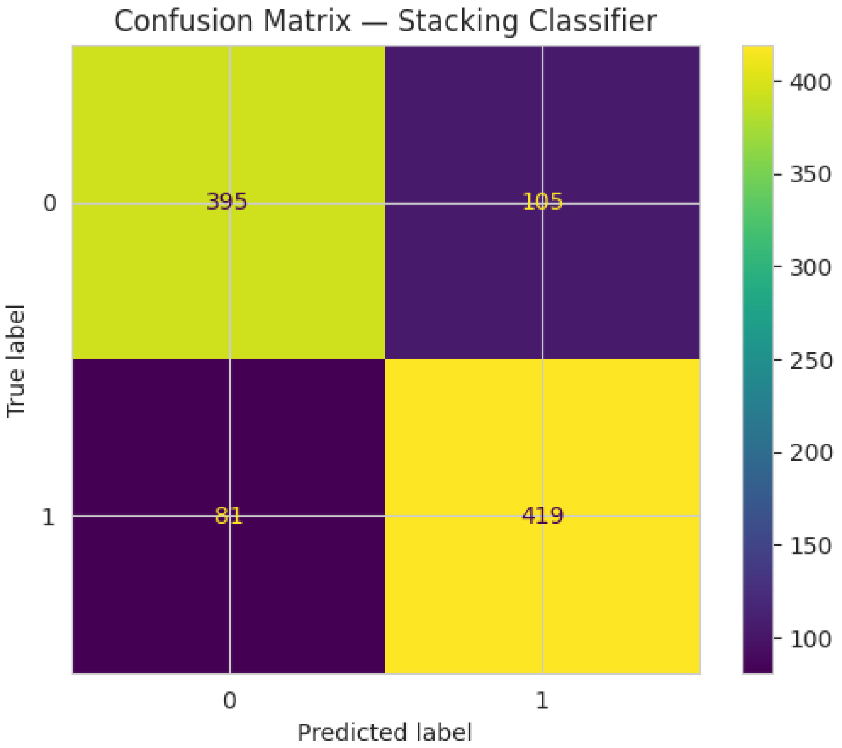
Models Trained

- Baseline models: Logistic Regression, SVM, KNN, RF
- Advanced models: XGBoost, MLP
- Ensemble methods: Voting, Stacking
- Best (MCYT): MLP & RF→ F1 > 0.84, EER ~ 0.16



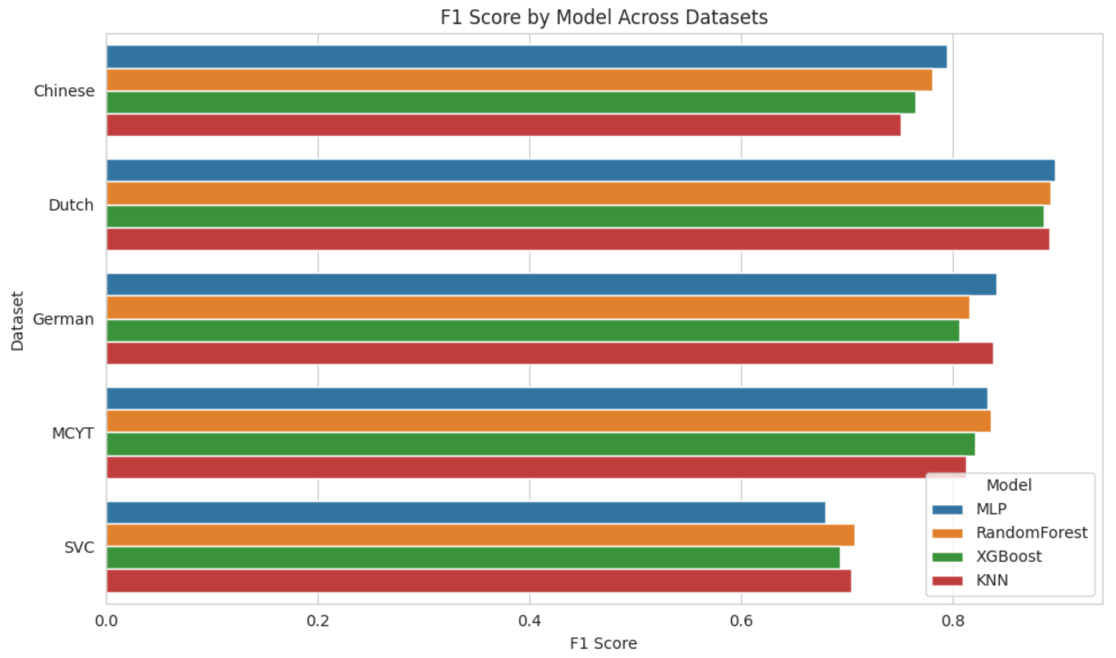
Ensemble Performance

- VotingClassifier: best balance of precision/recall
- StackingClassifier: didn't outperform simpler models

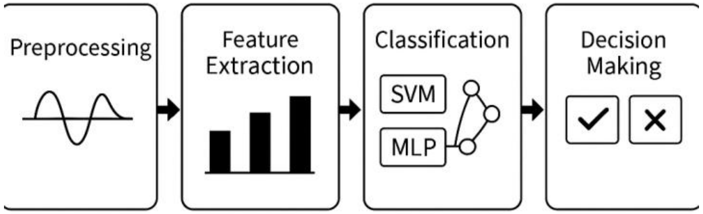


Cross-Dataset Comparison

- Dutch: easiest dataset (F1 > 0.89)
- SVC: hardest (F1 ~ 0.71, EER > 0.3)
- MLP most consistent across datasets



Key Takeaways



- X and Y were most influential (Random Forest)
- Complex models help, but not always better than RF
- Feature engineering was critical
- Ensemble models improve robustness

Conclusion & Future Work

- ML models can reliably detect forged signatures
- MLP & RF are ready for deployment

- **Future work:**

– Temporal sequence modeling (e.g., RNNs)

– Personalized thresholds per user

– Signature augmentation for low-data users