Emotion Detection from Keyboard Typing Patterns

Author: Souvik Swain

Date: 2025-08-15

# Abstract

This project explores the feasibility of predicting user emotional states from keystroke dynamics, using features such as key dwell time, inter-key latency, error/backspace rate, burstiness, and pauses. We train Random Forest and MLP baselines on session-level features and evaluate accuracy and class-wise F1.

# 1. Introduction & Motivation

Keystroke dynamics provide a low-cost, privacy-preserving signal about motor and cognitive status. Unlike text-based sentiment, this approach avoids storing raw content. Potential applications include adaptive UI, wellness nudges, or assisting accessibility.

# 2. Literature Snapshot

Prior work indicates that keystroke timing patterns can encode stress, workload, and affect. We extend this by using robust session-level features and discussing privacy-first design.

# 3. Dataset

We include a data template for per-keystroke logging and a synthetic session-level dataset for experimentation.

Template fields: user\_id, session\_id, utc\_timestamp\_iso, key, press\_duration\_ms, inter\_key\_delay\_ms, is\_error\_backspace, word\_len\_context, text\_domain, self\_report\_valence\_1to9, self\_report\_arousal\_1to9, label\_emotion

# 4. Methods

Feature engineering covers dwell/flight statistics (mean, std, CV), error/backspace rate, WPM, burstiness, pause ratios, and digraph variability. Models: RandomForest and MLPClassifier.

# 5. Experiments & Results

Using the synthetic dataset, both models achieve strong separation between emotions. Replace with results from your real data; include accuracy, macro-F1, confusion matrices, and per-class precision/recall.

# 6. Discussion

Generalization may vary across devices, languages, and contexts. Calibration per user can help. Future work: sequence models (LSTM/Transformer) on raw event streams; domain adaptation; active learning to reduce labeling burden.

# 7. Ethics & Privacy

Collect explicit consent, allow opt-out, encrypt at rest, and store only derived features where possible. Clarify that predictions are probabilistic and not medical diagnoses.

# 8. Conclusion

Keystroke dynamics are a promising modality for lightweight, privacy-aware emotion inference.

# Appendix: Reproducibility

Environment: Python 3.x, scikit-learn, pandas, numpy, matplotlib, joblib.

Code: see `train\_model.ipynb` and `collect\_typing\_data.py`.