Evaluating Machine Learning Approaches for Personality Recognition

Classifying Authors as Neurotic or Calm from Text

Introduction (1 of 1)

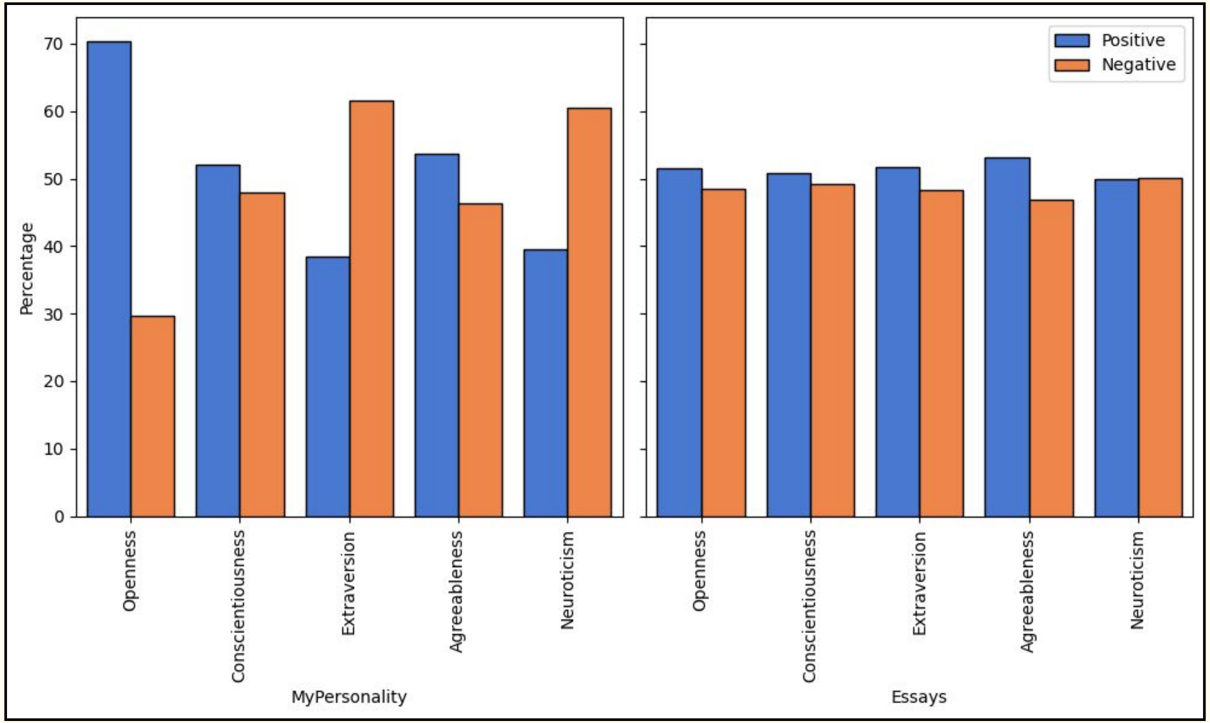
- Objective
  - Explore methods for classifying authors as "neurotic" or "calm" based on their writing
- Datasets
  - MyPersonality (train)
  - Essays (test)
- Methods
  - Evaluate different vectorizers (e.g., one-hot, term frequency, TF-IDF)
  - Evaluate different classifiers (e.g., SVM, k-NN, random forest, gradient boosting)
  - Grid search to optimize parameters

Background (1 of 2)

Label	Sample
	MyPersonality
Calm	Was stuck between reality and a dream...unpleasant. Let's go to Italia~~ Is going back to the homeland "Oh my God dude, doing 5 tomorrow is going to feel like sex!"
Neurotic	Just bought a cute pair of purple pumps :) Not yet - going to movies tonight... is sick in bed. Coughing like and old hag and looking decidedly ropey
	Essays
Calm	Well, I feel good about the fact that I am getting this assignment done well before it is due. Today is one of those days that I feel really motivated to do my homework
Neurotic	Well, right now I just woke up from a mid-day nap. It's sort of weird, but ever since I moved to Texas, I have had problems concentrating on things

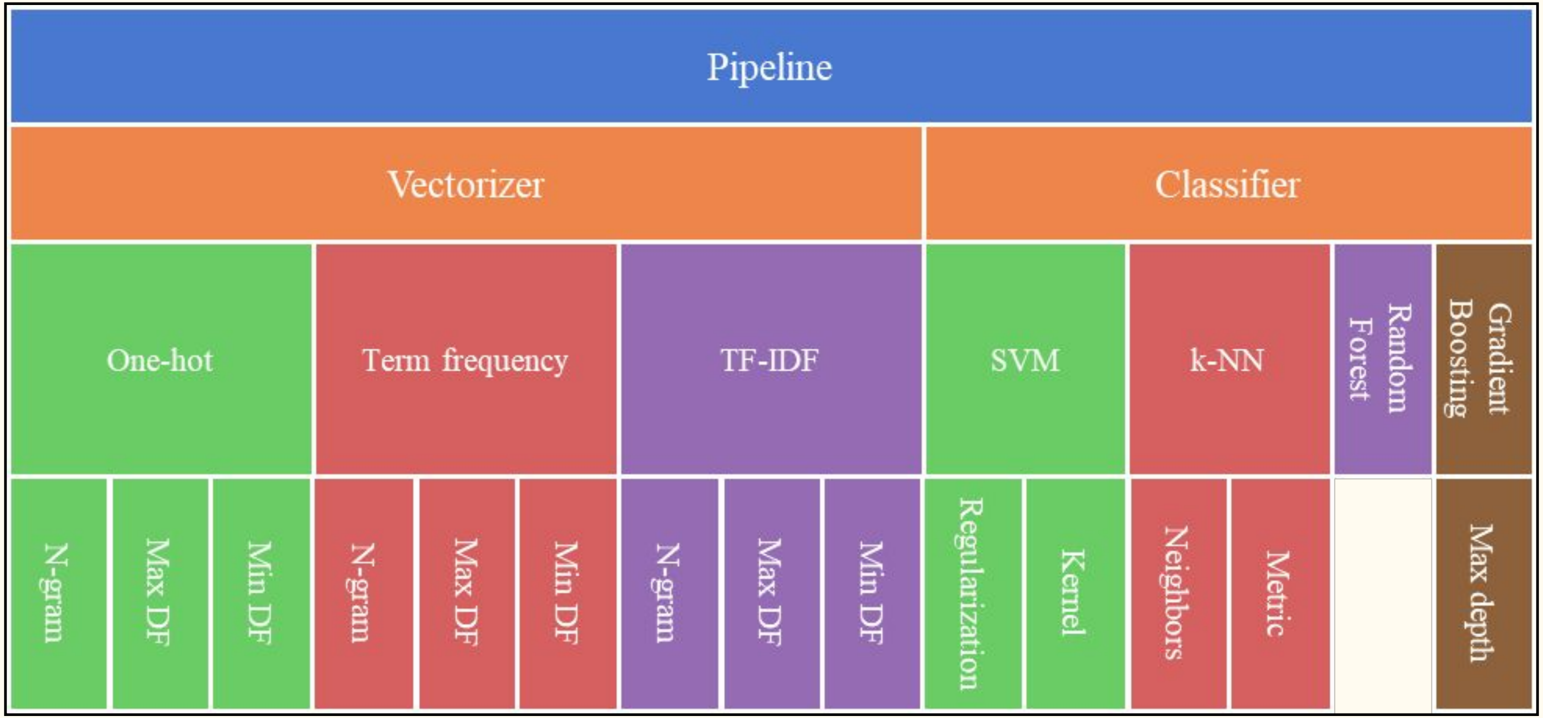
Background (2 of 2)

- Openness
  - Insightful
  - Unimaginative
- Conscientiousness
  - Organized
  - Careless
- Extraversion
  - Sociable
  - Shy
- Agreeableness
  - Friendly
  - Uncooperative
- Neuroticism
  - Neurotic
  - Calm



Author Distributions by text

Methods (1 of 2)



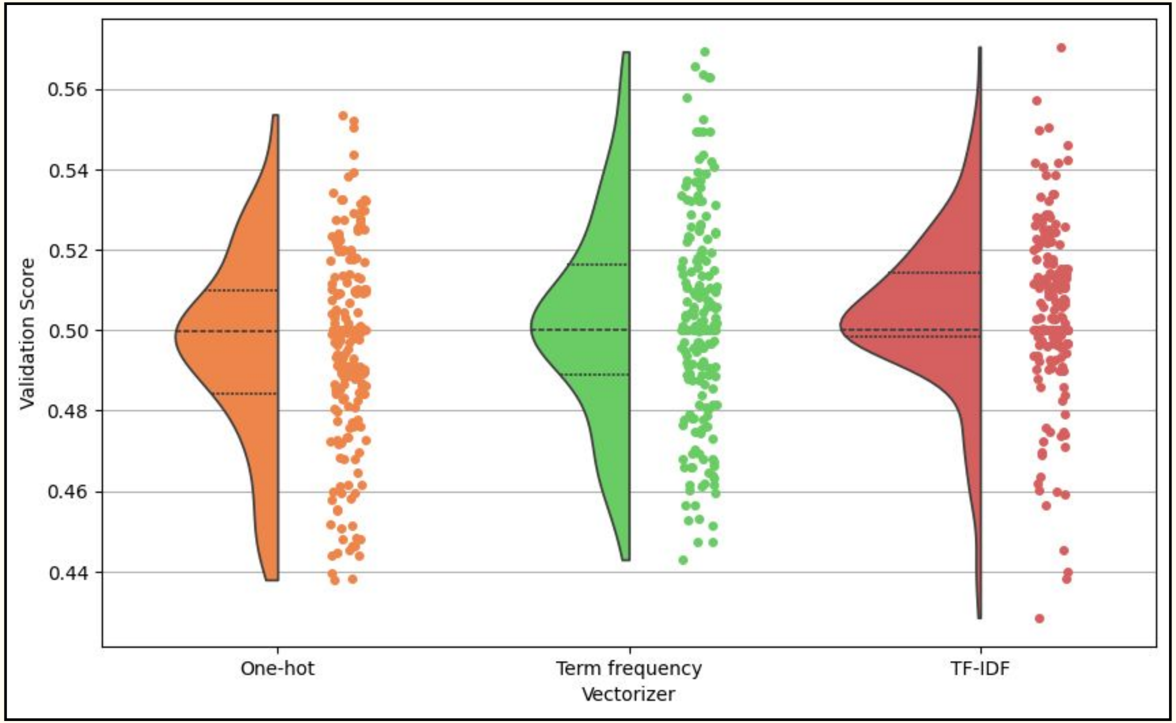
Methods (2 of 2)

- One-hot
  - N-gram (unigrams, bigrams)
  - Max DF (1.0, 0.5)
  - Min DF (1, 2)
- Term frequency
  - N-gram (unigrams, bigrams)
  - Max DF (1.0, 0.5)
  - Min DF (1, 2)
- TF-IDF
  - N-gram (unigrams, bigrams)
  - Max DF (1.0, 0.5)
  - Min DF (1, 2)

- SVM
  - Regularization (0.01, 0.1, 1, 10, 100)
  - Kernel (linear, RBF)
- k-NN
  - Neighbors (1, 3, 5, 7, 9)
  - Metric (Euclidean, cosine)
- Random forest
  - None :-)
- Gradient boosting
  - Max depth (1, 2, 3, 4, 5)

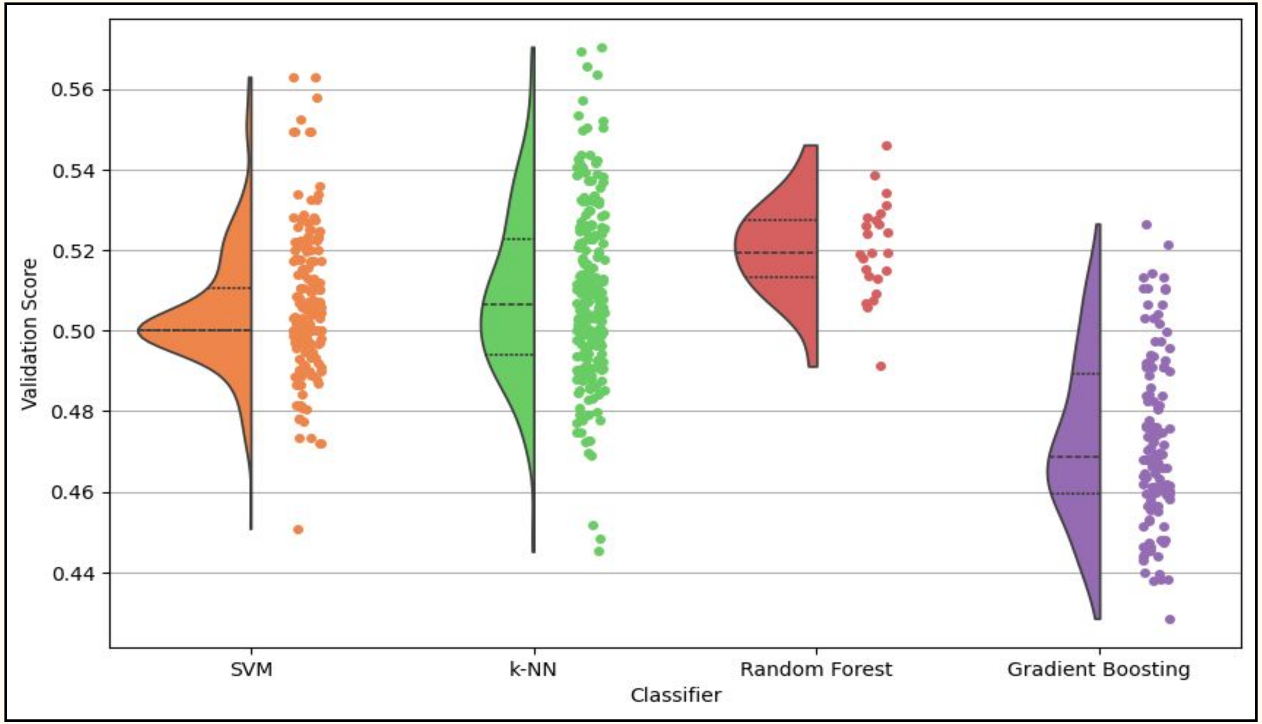
Results (1 of 3)

- One-hot
  - 0.50 ± 0.02 SD
  - 0.48–0.51 IQR
- Term frequency
  - 0.50 ± 0.03 SD
  - 0.48–0.52 IQR
- TF-IDF
  - 0.50 ± 0.02 SD
  - 0.50–0.51 IQR



Results (2 of 3)

- SVM
  - 0.50 ± 0.02 SD
  - 0.50–0.51 IQR
- k-NN
  - 0.51 ± 0.02 SD
  - 0.49–0.52 IQR
- Random forest
  - 0.52 ± 0.01 SD
  - 0.51–0.53 IQR
- Gradient boosting
  - 0.47 ± 0.02 SD
  - 0.46–0.49 IQR



Results (3 of 3)

TF-IDF Vectorizer			k-NN Classifier	
N-gram	Max DF	Min DF	k	Metric
Bigrams	0.5	2	9	Euclidean

Label	Precision	Recall	F1-Score	Support
Calm	0.51	0.47	0.49	1235
Neurotic	0.51	0.55	0.53	1233
Accuracy			0.51	2468
Macro Average	0.51	0.51	0.51	2468
Weighted Average	0.51	0.51	0.51	2468

Discussion (1 of 2)

- SVM
  - Consistent results with term frequency vectorizer (unigrams/bigrams)
  - Best with linear kernel, and regularization strength of 10
- k-NN
  - Best overall performance with TF-IDF or term frequency
  - Optimal with 9 neighbors (TF-IDF) and 3 neighbors (term frequency)
- Random Forest
  - Stable but underperformed compared to k-NN and SVM
  - TF-IDF with bigrams slightly improved results
- Gradient Boosting
  - Higher variability and lower performance than other models

Discussion (2 of 2)

- Comparison
  - Mohammad et al. got an accuracy of 0.57 with SVM and external resources
  - Iacobelli et al. got an accuracy of 0.56 using Naïve Bayes
  - Markovic et al. got an F1-score of 0.90 with SVM and boosting, using a larger feature set
- Limitations
  - LIWC, MRC, or Twitter Hashtag Lexicon could improve analysis
  - A larger, balanced dataset would improve performance
  - Adding demographics or other features could yield richer patterns
  - Word embeddings or deep learning could improve semantic understanding

Conclusion (1 of 1)

- Objective
  - Explore machine learning models for personality recognition from text
- Methods
  - Grid search of vectorizers (one-hot, term frequency, TF-IDF) and classifiers (SVM, k-NN, random forest, gradient boosting)
- Key Findings
  - k-NN with TF-IDF achieved competitive results with an F1-score of 0.51
- Improvement
  - Incorporating external resources, increasing feature space, and using larger dataset could enhance accuracy
- Future Work
  - Word embeddings, deep learning, and multimodal data could improve model performance and semantic understanding

References

Celli, F., Pianesi, F., Stillwell, D., & Kosinski, M. (2013). Workshop on computational personality recognition: Shared task. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 7, No. 2, pp. 2-5).