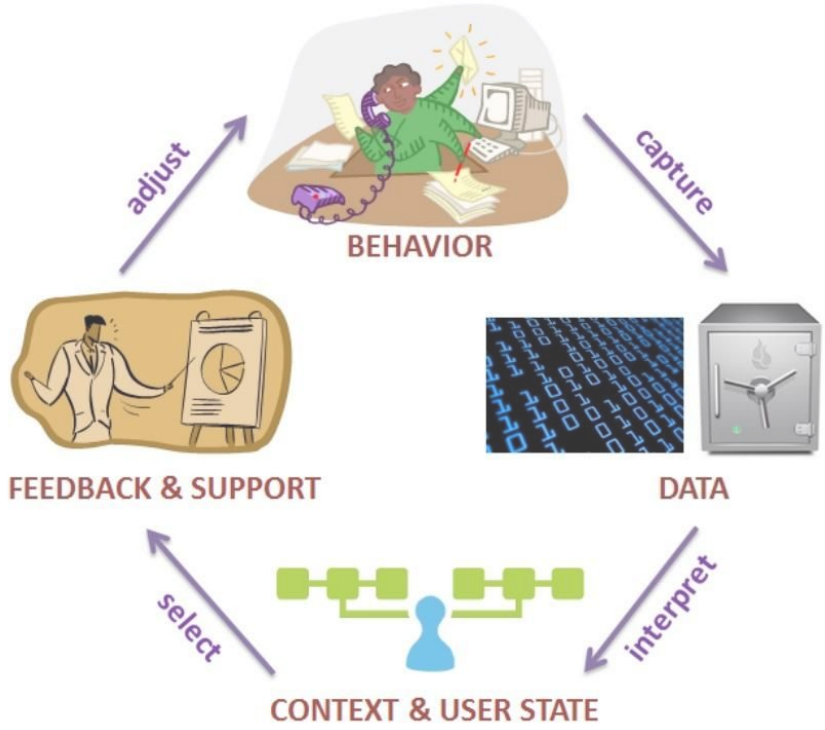
Investigating Supervised Classification on Stress Recognition: The Case of SWELL - KW Dataset

Dilruba Showkat

May 6, 2022

Overview

- ❑ Dataset Description and Objective
- ❑ Methods
- ❑ Results and Discussion



Dataset Description

Dataset: Smart Reasoning for Well-being at Home and at Work (SWELL- KW)

- Created from an experiment with 25 people doing knowledge work (e.g., searching for info)

- Train
  → Sample size: 369289
  → Total features: 36 numerical attributes
  → Labels: class 1-3

- Test
  → Training Sample size: 41033
  → Test features size: 36
  → Labels: class 1-3

- Samples Per Class (Training Set)
  → No-stress: 200082
  → Interruption: 105150
  → Time pressure: 64057

Saskia Koldijk, Maya Sappelli, Suzan Verberne, Mark A Neerincx, and Wessel Kraaij. 2014. The swell knowledge work dataset for stress and user modeling research. In Proceedings of the 16th international conference on multimodal interaction. 291–298

Objective

- **Primary Objective:** to investigate various supervised classification algorithms on the dataset.

Methods

Data Preprocessing

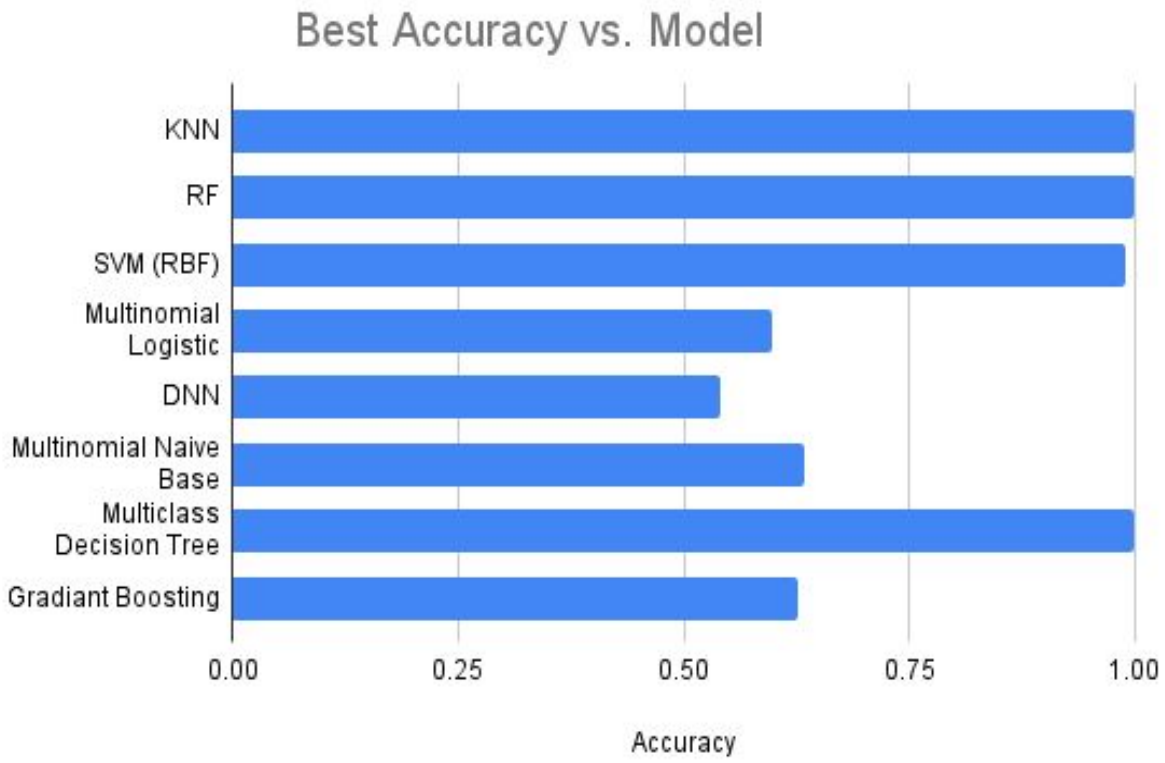
- Class label encoding; drop sampen, higuci, datasetId
- Cross Validation: K fold
- Feature Engineering:
  - ‘Feature importance’ and ‘select K best’ method
    - Select top 5 scored subset of features: MEAN_RR, MEDIAN_RR, pNN25, pNN50, HR
  - PCA
    - n_components = 5, 10

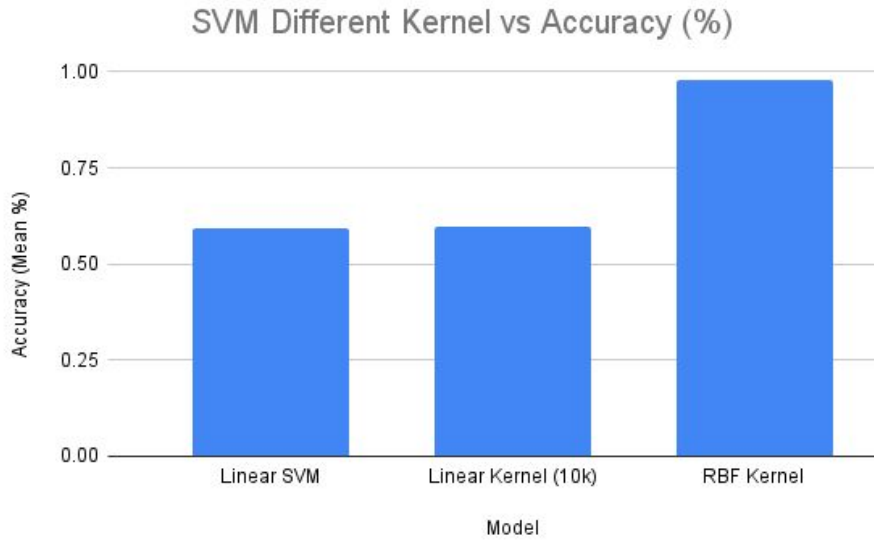
Algorithms

- K-Nearest Neighbor (KNN)
- Support Vector Machine (SVM)
- Multi-class Decision Tree (DT)
- Multi-class Random Forest (RF)
- Deep Neural Network (DNN) with Pytorch
- Multinomial Naive Bayes (NB)/ Gaussian NB
- Multi-class Logistic Regression (LR)

Results

Results: Best Performance Across Models

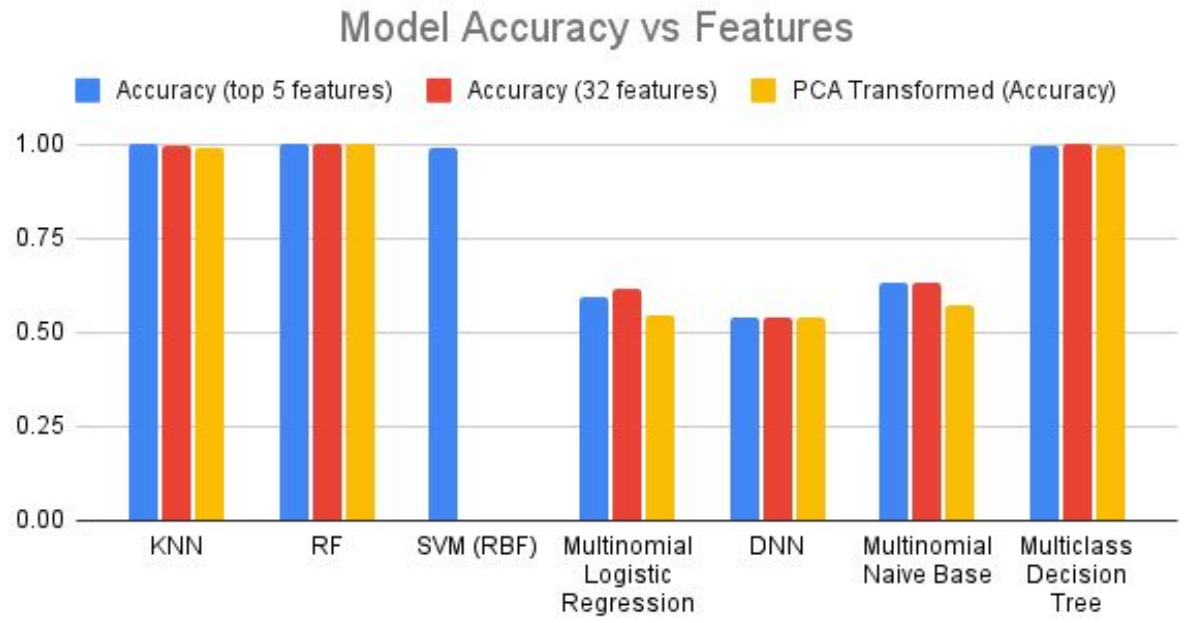




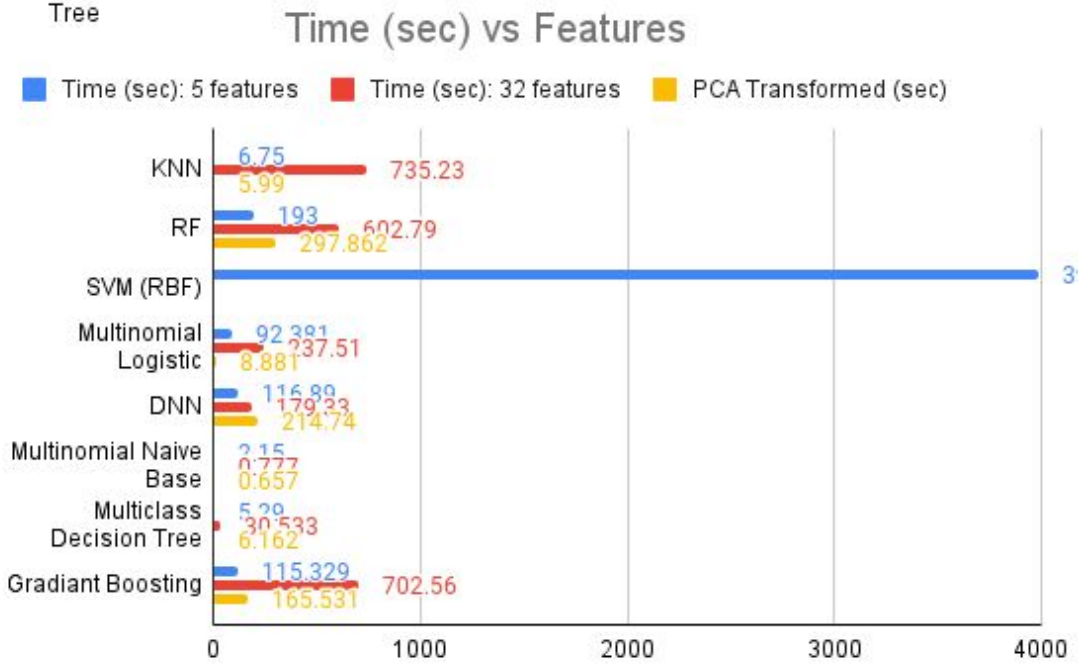
Results: Optimal Parameter Set Across Models

Model	Optimal Parameter Set	Accuracy	Time (sec)
KNN	k = 5	.99	5.99
RF	criterion = gini, n_estimator = 200, max_depth = 30	1.0	193
SVM	kernel = rbf, $\gamma = .01, C = 50$	.99	3986.188
Multi-LR	Regularizer = L2, C = 1000, solver = lbgfs	.59	92.3
NB	var smooting = .0018	.63	0.77
DT	depth = 100, criterion = gini	.948	5.29
DNN	num_layer = 3, hidden-layer size = 512, lr = 0.01 or 0.05	.54	1222.62

Results: Accuracy and Time Across Features Set

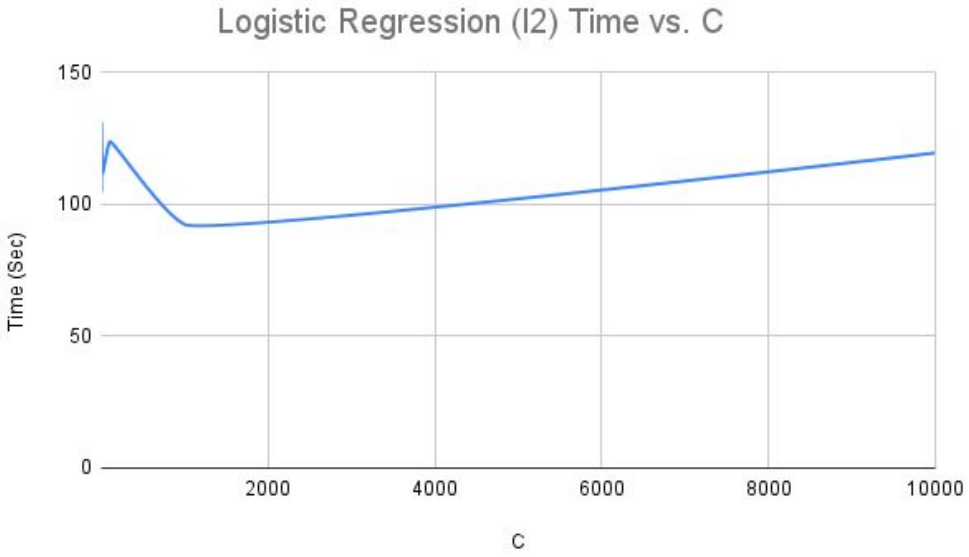


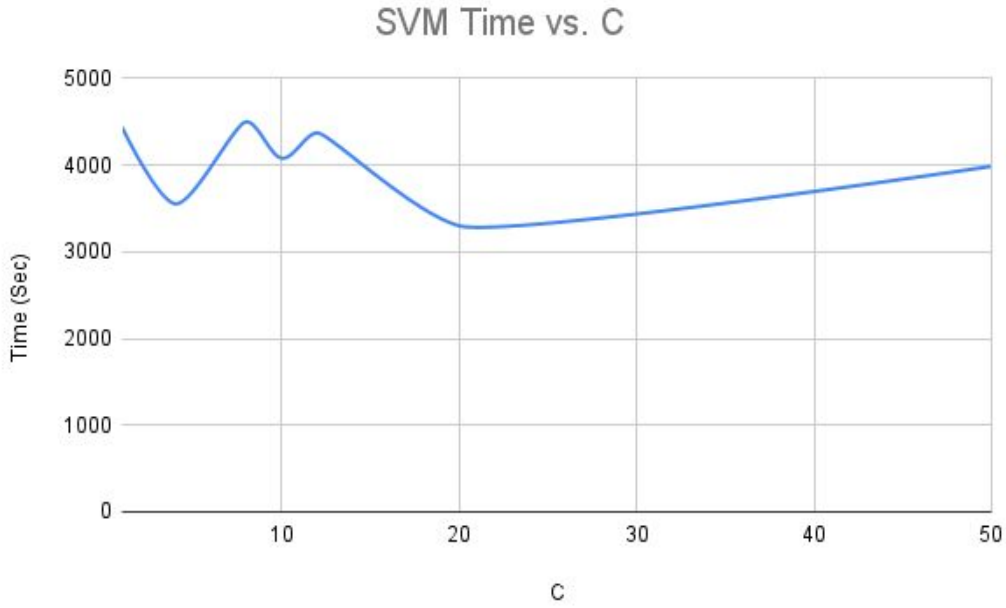
Accuracy was almost same across different feature set.

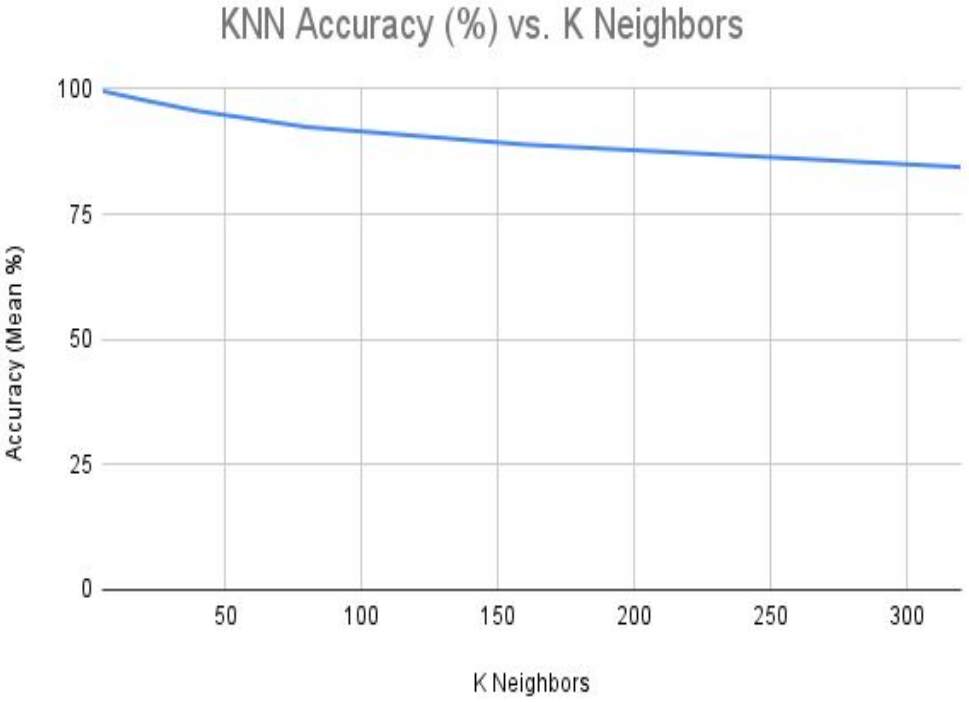


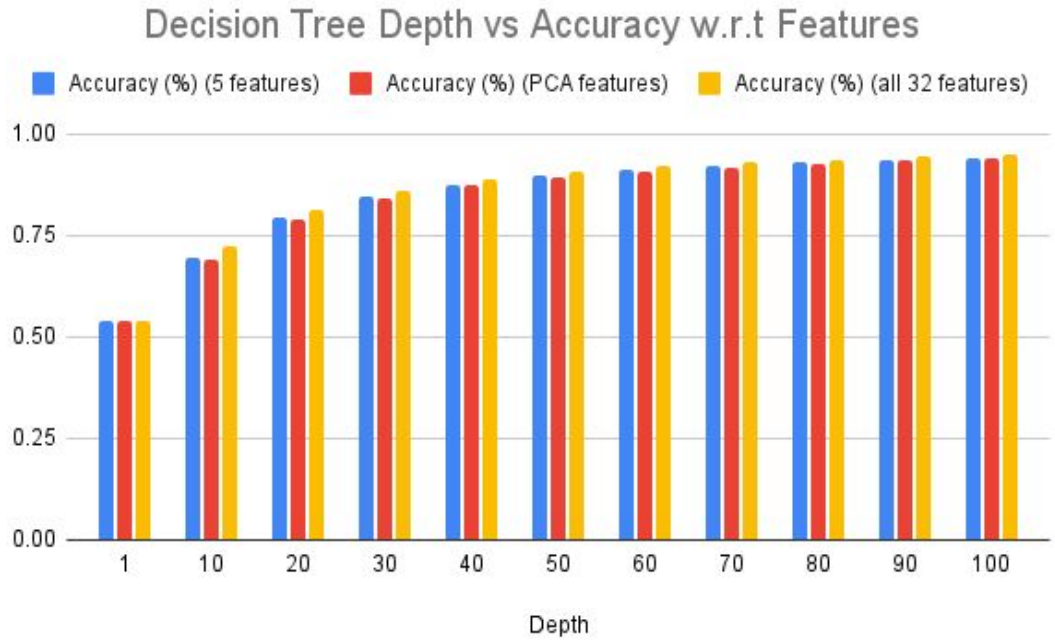
Running time increased with increasing number of features.

Results: Parameter Tuning Across Models (1)









Results: Parameter Tuning for DNN (2)

Train Batch	LR	Train	Validation	Test	Time (sec)
20	0.05	28.47	28	28	7874.09
32	0.1	54.19	54	54	8416.52
64	0.01	54.17	54	54	8734.35
128	0.001	68.61	75	44	8947.85
512	0.0001	59.84	72	42	9082.19

- Increasing **batch size** and decreasing ***lr*** improved train/validation score, but not test score.
- Increasing ***hidden layer*** and ***neuron size*** didn't improve test accuracy for NN, however, they significantly increased running time

Key Takeaways

- In addition to exhaustive hyper-parameter tuning and experimentation, this work
  - improved accuracy over several related prior work and
  - propose 5 optimal set of parameter {$MEAN\_RR$, $MEDIAN\_RR$, $pNN25$, $pNN50$, HR} as opposed to 10 in prior work without compromising accuracy or running time.

[1] Saskia Koldijk, Maya Sappelli, Suzan Verberne, Mark A Neerincx, and Wessel Kraaij. 2014. The swell knowledge work dataset for stress and user modeling research. In Proceedings of the 16th international conference on multimodal interaction. 291–298

[2] Chang Su. 2020. Heart Rate Variability Feature Selection using Random Forest for Mental Stress Quantification. Ph.D. Dissertation. Concordia University

Thank you!