Kushagra Bhatia\* and Arpit Sharma

\*kushagra1198@gmail.com

Sector classification for crowd-based software Requirements

TABLE OF CONTENTS

Dataset 2-3

Sample Requirements 2

Descriptive statistics 2

Sector-wise most frequent words 2

Additional stop-word list 2

Word clouds for each sector 3

Approach Implementation 4-6

Approach 1 4-5

Approach 2 5

Approach 3 5-6

Experimental evaluation 7-8

Results for all three approaches. 7

Sector-wise Accuracy of the best model in each approach 8

# Dataset

1. Sample Requirements

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Role** | **Feature** | **Benefit** | **Application Domain** | **Application domain\_other** | **Tags** |
| worker | my smart home to be able to order delivery food by simple voice command | I can prepare dinner easily after a long day at work | Health |  | food, delivery, dinner, voice |
| home occupant | my smart home to turn on certain lights at dusk | I can come home to a well-lit house | Energy |  | lights, turn on, night |
| worker | my smart home to sync with my biorhythm app and turn on some music that might suit my mood when I arrive home from work | I can be relaxed | Entertainment |  | music, biorhythm, mood |
| parent | my smart home to keep me up to date about my children's activities when I'm out of the home | I can know they're safe and positively occupied | Safety |  | child, monitor, status |
| home occupant | my smart home to sync all Christmas-related lights, indoor and outdoor, and turn them on and off at the appropriate time | it's more convenient for me | Other | Convenience | lights, turn on, turn off, indoor, outdoor, Christmas, synchronize |

1. Descriptive Statistics of Dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Application Domain | Requirements in Train-Set | Requirements in Test-Set | Total Requirements |
| Energy | 532 | 94 | 626 |
| Entertainment | 400 | 71 | 471 |
| Health | 504 | 89 | 593 |
| Safety | 758 | 134 | 892 |
| Other | 327 | 57 | 384 |

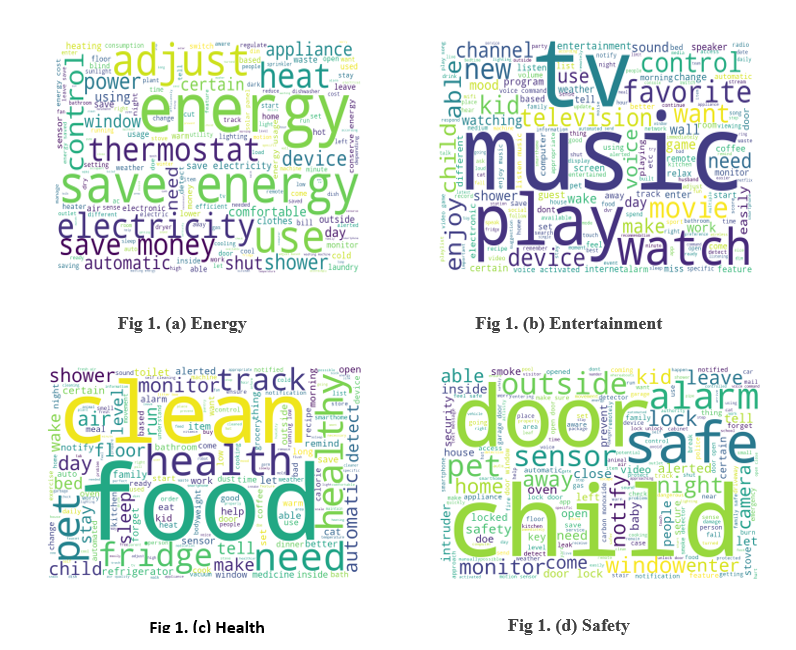
1. Sector-wise most frequent words

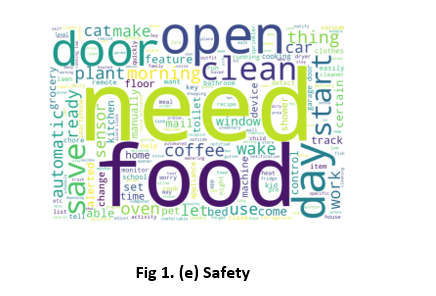
|  |  |
| --- | --- |
| Domain | Top 14 most frequent words |
| Energy | home, smart, food, time, know, clean, automatically, house, water, temperature, air, health, need, pet |
| Entertainment | energy, home, save, smart, turn, room, light, water, temperature, automatically, electricity, house, time, money |
| Health | home, music, smart, room, time, tv, play, turn, house, voice, automatically, watch, favorite, movie |
| Safety | home, door, smart, house, know, alert, safe, child, lock, alarm, automatically, open, window, pet |
| Other | home, smart, time, automatically, door, water, know, house, need, open, food, dog, day, turn |

1. Additional stop-word list:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| home | smart | time | automatically | house | know | water |
| turn | temperature | shower | alert | dog | room | phone |

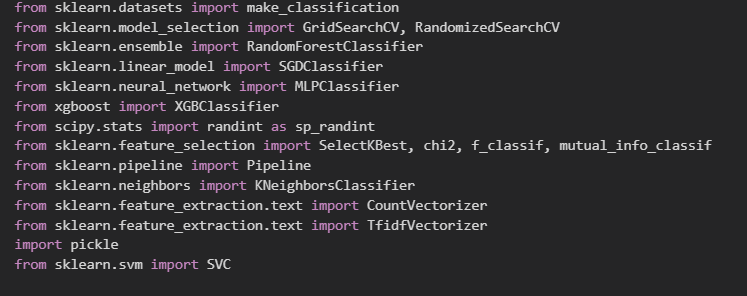
1. Word clouds for each domain (after removal of stop-words and additional stop-words)



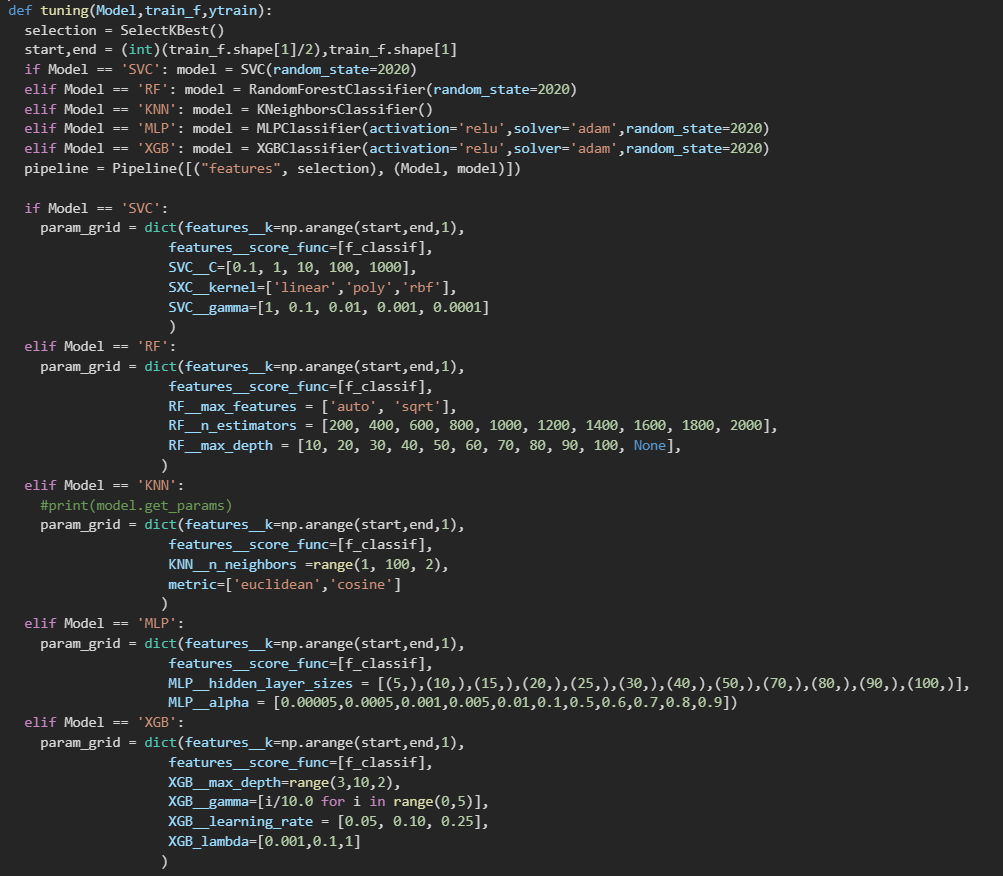


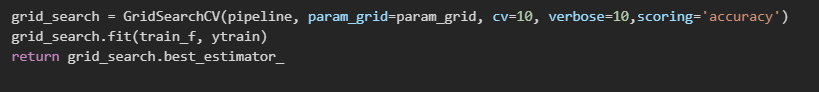
# Implementation

1. Approach 1 (Implemented with Sklearn)
2. Importing required libraries.



1. Define grid to be searched over on the pipeline.

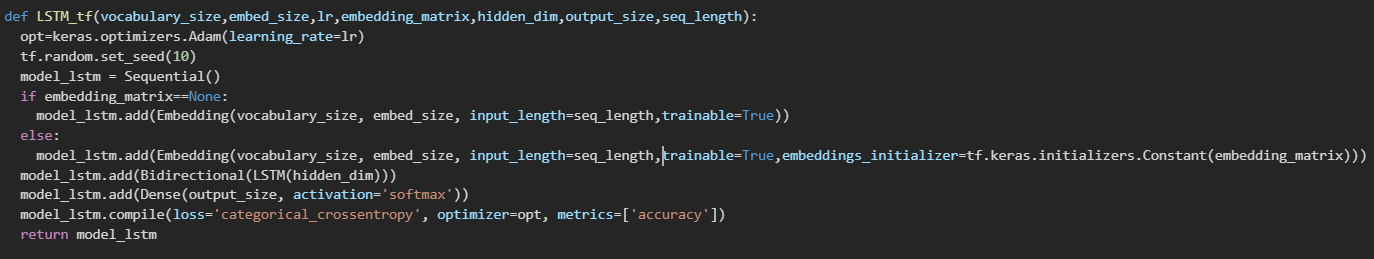




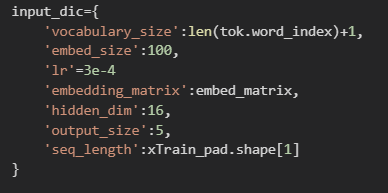
1. Feature Importance for Hand-engineered features

|  |  |
| --- | --- |
| **Feature** | **ANOVA F-value** |
| length of requirement in characters | 0.32879427 |
| percentage of hyperbolic words | 0.40527909 |
| percentage of contractions | 0.47148293 |
| average word size | 5.19404783 |
| percentage of stop words | 5.69315744 |
| maximum word size | 7.62893018 |
| percentage of easy words | 8.11528589 |

1. Approach 2 (Implemented with Keras)
2. LSTM model architecture used in approach 2, with Adam optimizer.



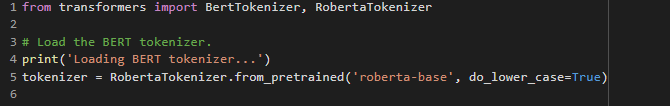
1. Selected hyperparameters after grid search (with pre-trained glove embedding)



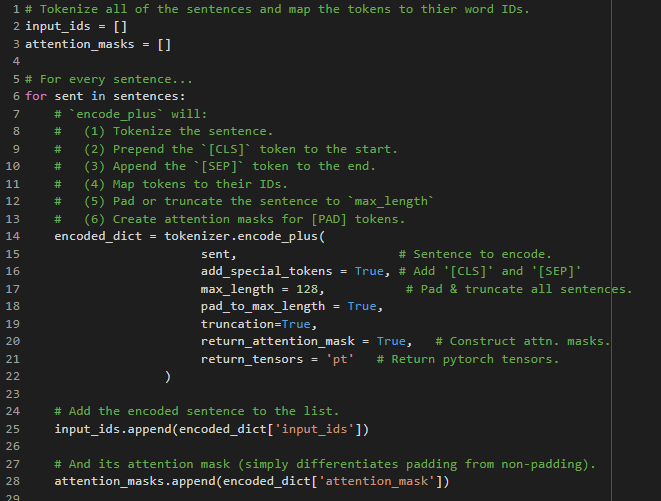
Note: 1. In case of random initialization we do not argument an embedding matrix.

2. We use the [1] architecture for CNN implementation.

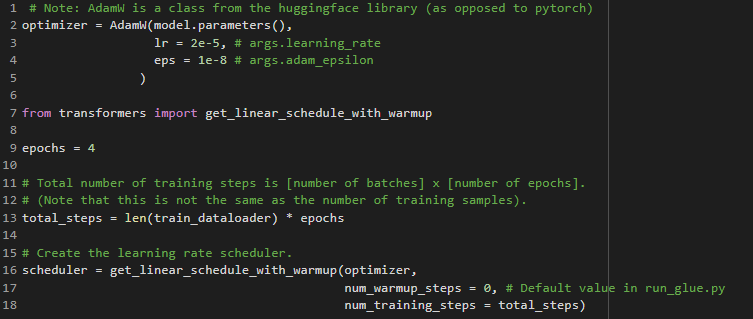
1. Approach 3 (Implemented with Pytorch)
2. Loading Huggingface’s implementation



1. Processing the requirements before fine-tuning the transformer model



1. Optimizer and scheduler for training



# Experimental Evaluation

1. Results for all three approaches.
2. Approach 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Sentence Encoding** | **k** | **Precision** | **Recall** | **Fscore** |
| Random Forest | Averaged Glove + features | 101 | 64.29 | 67.19 | 63.21 |
| TF-IDF Weighted Glove + features | 95 | 61.6 | 64.26 | 62.05 |
| Universal Sentence Encoding | 330 | **70.01** | **71.68** | 69.71 |
| XGBoost | Averaged Glove + features | 93 | 65.15 | 66.74 | 65.63 |
| TF-IDF Weighted Glove + features | 104 | 65.07 | 67.19 | 65.48 |
| Universal Sentence Encoding | 341 | 68.64 | 70.33 | 68.87 |
| SVM | Averaged Glove + features | 83 | 68.3 | 70.78 | 69.11 |
| TF-IDF Weighted Glove + features | 101 | 67.71 | 69.88 | 68.2 |
| Universal Sentence Encoding | 519 | 68.42 | 71.46 | 69.44 |
| KNN | Averaged Glove + features | 59 | 63.8 | 66.29 | 63.96 |
| TF-IDF Weighted Glove + features | 72 | 62.52 | 65.39 | 61.73 |
| Universal Sentence Encoding | 384 | 67.99 | 70.11 | 67.8 |
| MLP | Averaged Glove + features | 59 | 66.97 | 68.31 | 67.33 |
| TF-IDF Weighted Glove + features | 72 | 63.34 | 65.39 | 63.3 |
| Universal Sentence Encoding | 519 | 69.9 | 71.46 | **70.46** |

1. Approach 2

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Precision** | **Recall** | **F-score** |
| CNN | 66.45 | 68.31` | 65.07 |
| CNN+P.T Glove | 68.77 | 70.11 | 67.75 |
| Bi-LSTM | 68.02 | 70.33 | 68.54 |
| Bi-LSTM+P.T Glove | **71.39** | **72.36** | **70.81** |

1. Approach 3

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Precision** | **Recall** | **F-score** |
| BERT-base-uncased | 72.25 | 74.15 | 72.14 |
| DistilRoBERTa-base | 73.69 | 75.06 | 73.55 |
| RoBERTa-base | **75.56** | **75.73** | **75.13** |

1. Sector-wise Accuracy of the best model in each approach

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Energy | Entertainment | Health | Safety | Other |
| RF+USE | 75.53 | 78.87 | 79.77 | 82.83 | 17.54 |
| LSTM+P.T Glove | 76.59 | 78.87 | 75.20 | 85.07 | 22.80 |
| RoBERTa-base | 78.72 | 80.28 | 78.65 | 85.07 | 38.59 |

References

[1] Yoon Kim. 2014. Convolutional neural networks for sentence classification. arXiv

preprint arXiv:1408.5882 (2014).