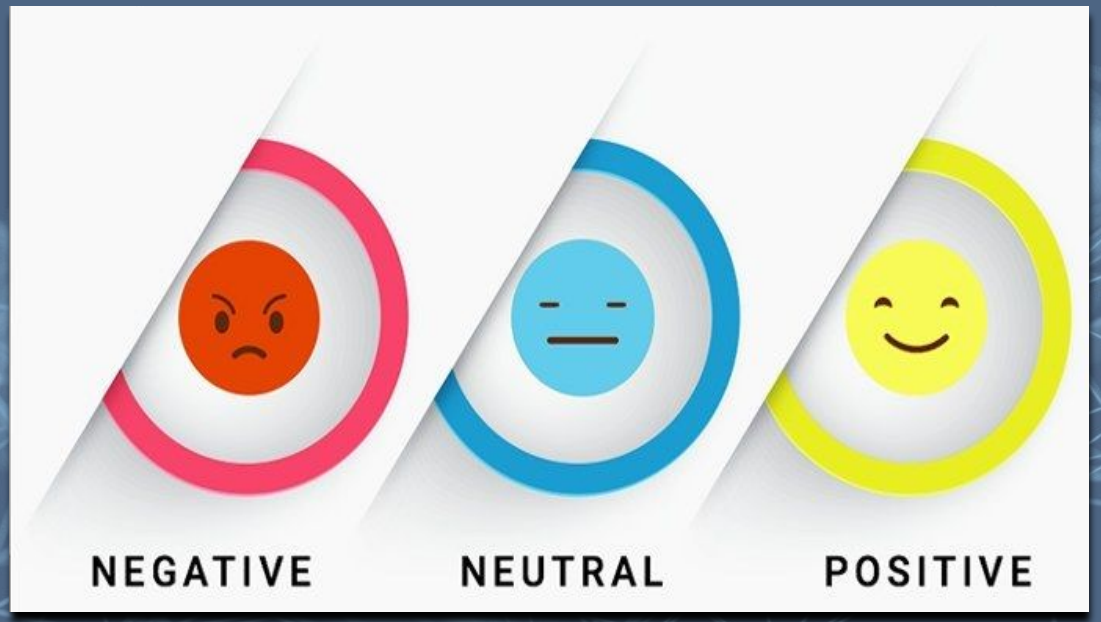
Sentiment Detection



The Team

Title	Name
Researcher	Loyiso Jiya
Co-researcher	Dayaan Sataar
Supervisor	Mehrdad Ghaziasgar
Co-supervisor	Reginald Dodds

Background

- **What is sentiment detection?**
  - Attempts to determine the feeling conveyed by any text
  - Can be positive, neutral, negative, or range
  - Text can be documents, forum posts or posts on social media
  - AKA sentiment analysis, opinion mining, opinion extraction
- **Why sentiment detection?**
  - Companies need feedback
  - conducted market research, focus groups, etc.
  - Inefficiency, time-consuming and costly
  - reviews after service is issued (i.e. MTN customer service)
  - reviews when given are often bias (mostly negative)

More Background

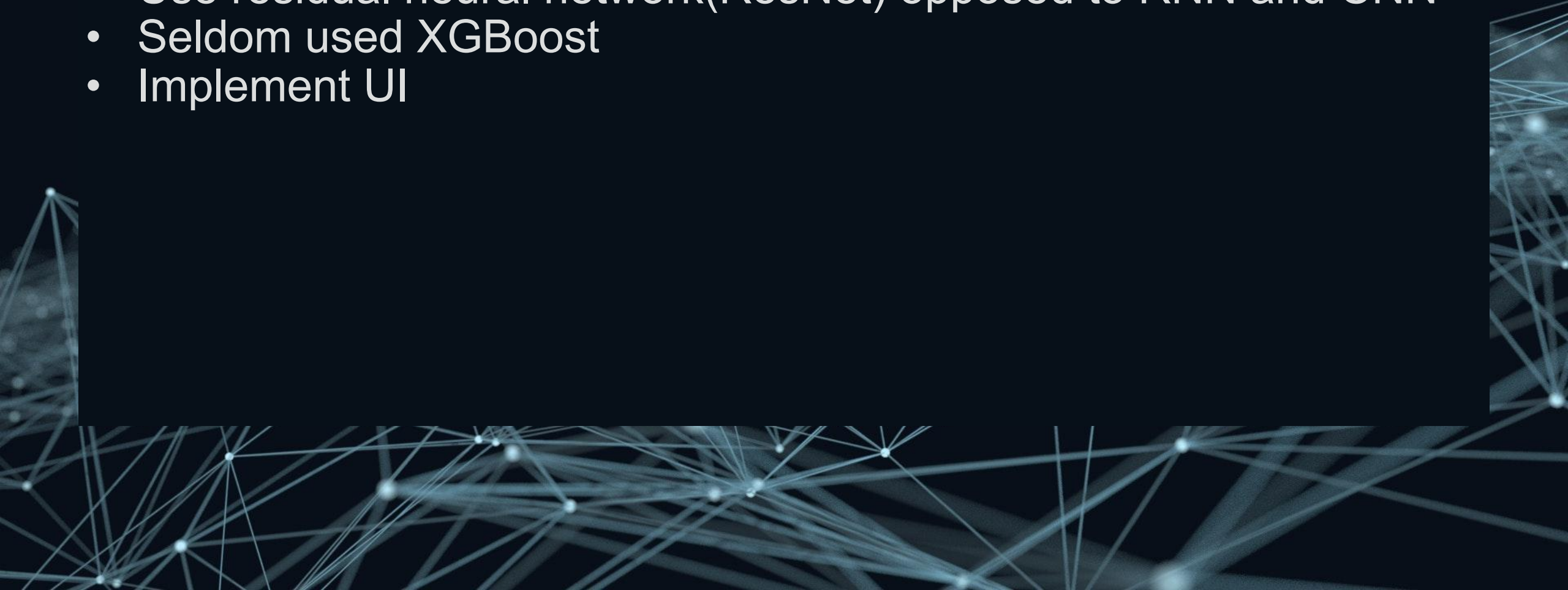
- Exponential growth of social media means free data
- Unbiased, unfiltered, and subjective to users
- Praise or criticisms of companies, political and social commentary etc.
- Gather and analyse data on a large scale
- Of these Twitter has about 300 mil users per month [1]
- Twitter is more favourable
- Twitter has open API
- Dell and Amazon use Twitter sentiment detection[2]
- Locally, Dr Ghaziasgar supervised an employee of one of the telecoms companies

Related Work

Authors	Dataset	Pre-processing	Technique	Feature Selection	Feature Extraction
Fouad et al. [3]	STS, Sanders and HCR	Tokenizing, removing links and emoticons	SVM, Linear regression, Naïve Bayes, and Majority Voting Ensemble	Information Gain	BoW, PoS tagging
Zheng et al. [4]	Online reviews from Chinese ecommerce website	None mentioned	SVM	None Mentioned	TF-IDF N-PoS-Grams
Jiaqiang and Xiaolin[5]	STS, SemEval2014, SS-Twitter, SE-Twitter, STS-Gold	None Mentioned	Random Forests, Naïve Bayes, SVM and Logistic regression	None Mentioned	Prior polarity, N-grams
Al-Smadi et al. [6]	SemanticEavaluation Workship 2016 (SemEval-ABSA16)	Remove unwanted special and Latin characters.	RNN and SVM	None mentioned	N-grams, PoS, Semantic features, word2vec
Zainuddin et al.[7]	STS and Sanders	Remove: hashtags, links, usernames, punctuation marks etc.	SVM	PCA, LSA and RP	PoS
Lauren et al. [8]	Sentiment Polarity Dataset v2.0	None mentioned	Logistic regression	PCA	ELM and word2vec

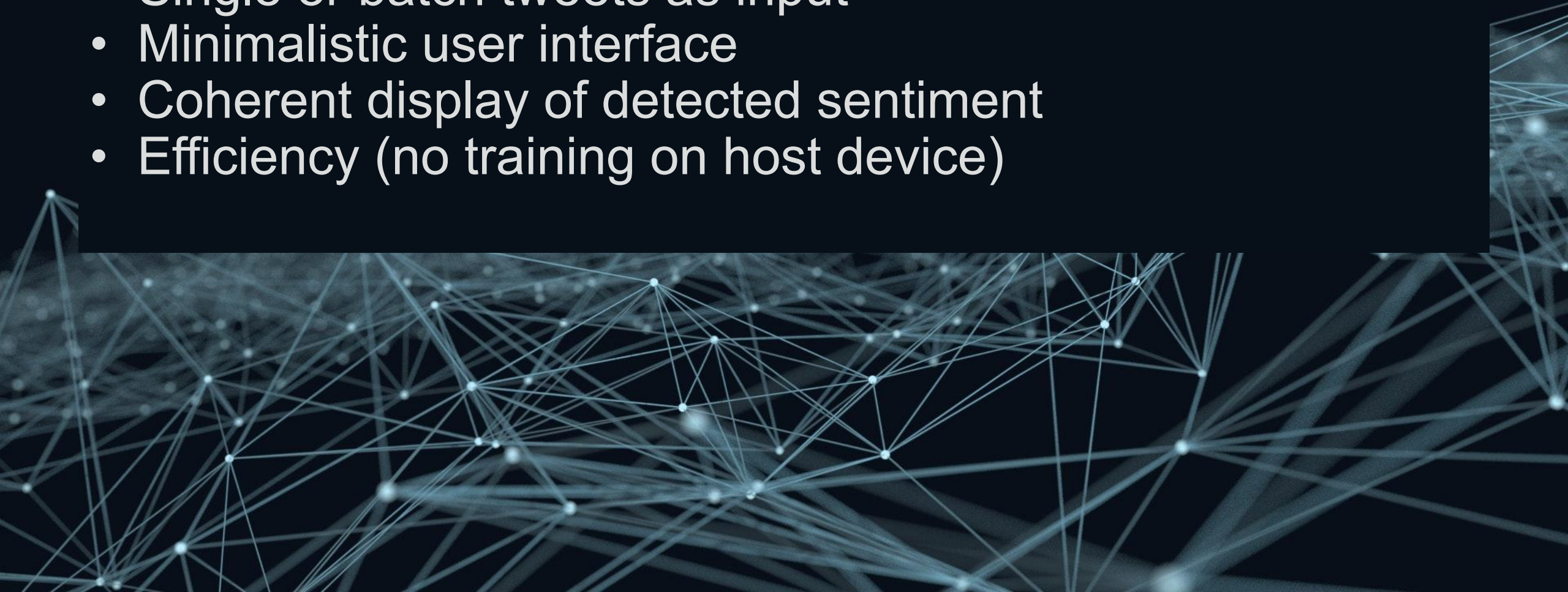
Our Contribution

- Compare feature extraction and feature selection techniques
- Use residual neural network(ResNet) opposed to RNN and CNN
- Seldom used XGBoost
- Implement UI



User Requirements

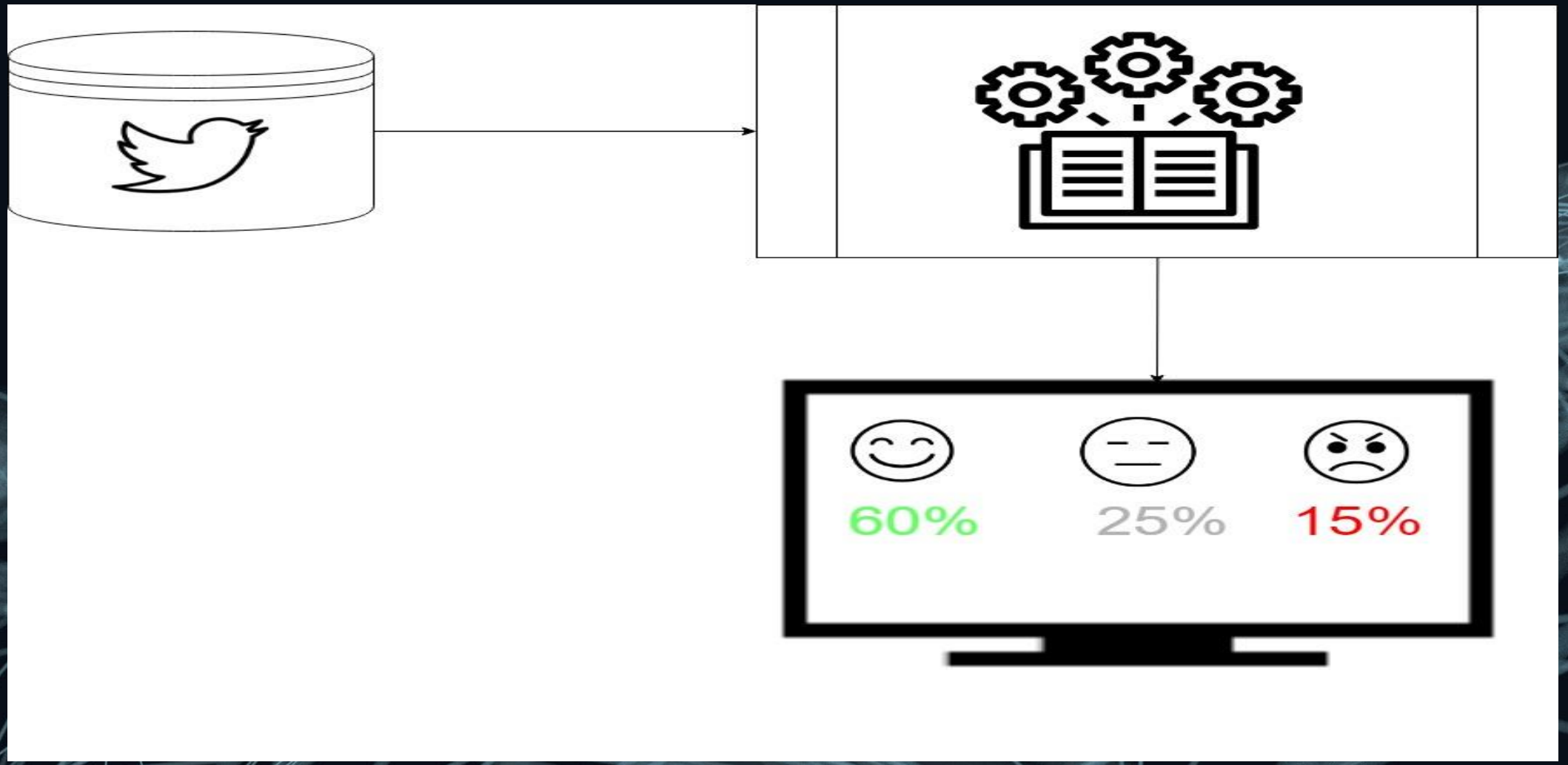
- From focus group within the research group
- Single or batch tweets as input
- Minimalistic user interface
- Coherent display of detected sentiment
- Efficiency (no training on host device)



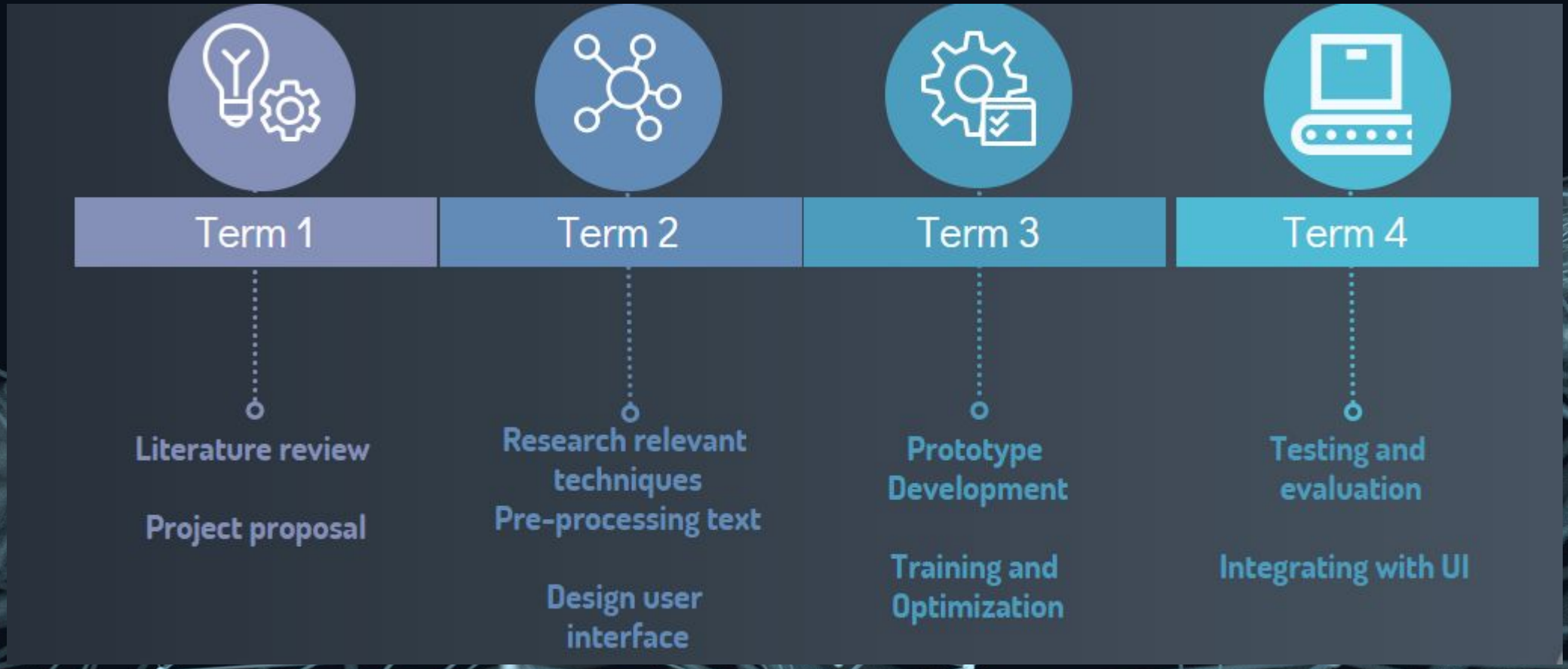
Requirement Analysis

- Stanford Twitter Corpus dataset
- Detect sentiment from given text (string or csv with multiple strings)
- Present detected sentiment to the user (graphically)

Visual Layout



Project Plan



References

1. Investor.twitterinc.com. Quarterly results. 2018. [Online]. Available: https://investor.twitterinc.com/financial-information/quarterly-results/default.asp
2. M. Oussalah, F. Bhat, K. Challis, and T. Schnier, “A software architecture for twitter collection, search and geolocation services,” Knowledge-Based Systems, vol. 37, pp. 105–120, 2013.
3. M. M. Fouad, T. F. Gharib, and A. S. Mashat, “Efficient twitter sentiment analysis system with feature selection and lassifier ensemble,” in International Conference on Advanced Machine Learning Technologies and Applications. Springer, 2018, pp. 516–527.
4. L. Zheng, H. Wang, and S. Gao, “Sentimental feature selection for sentiment analysis of Chinese online reviews,” International journal of machine learning and cybernetics, vol. 9, no. 1, pp. 75–84, 2018.
5. Z. Jianqiang and G. Xiaolin, “Comparison research on text preprocessing methods on twitter sentiment analysis,” IEEE Access, vol. 5, pp. 2870–2879, 2017.
6. M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, “Deep recurrent neural network vs. support vector machine for aspect based sentiment analysis of Arabic hotels reviews,” Journal of Computational Science, vol. 27, pp. 386–393, 2018.
7. N. Zainuddin, A. Selamat, and R. Ibrahim, “Hybrid sentiment classification on twitter aspect-based sentiment analysis,” Applied Intelligence, pp. 1–15, 2018.
8. P. Lauren, G. Qu, J. Yang, P. Watta, G.-B. Huang, and A. Lendasse,“Generating word embeddings from an extreme learning machine for sentiment analysis and sequence labelling tasks,” Cognitive Computation,vol. 10, no. 4, pp. 625–638, 2018.

Questions…