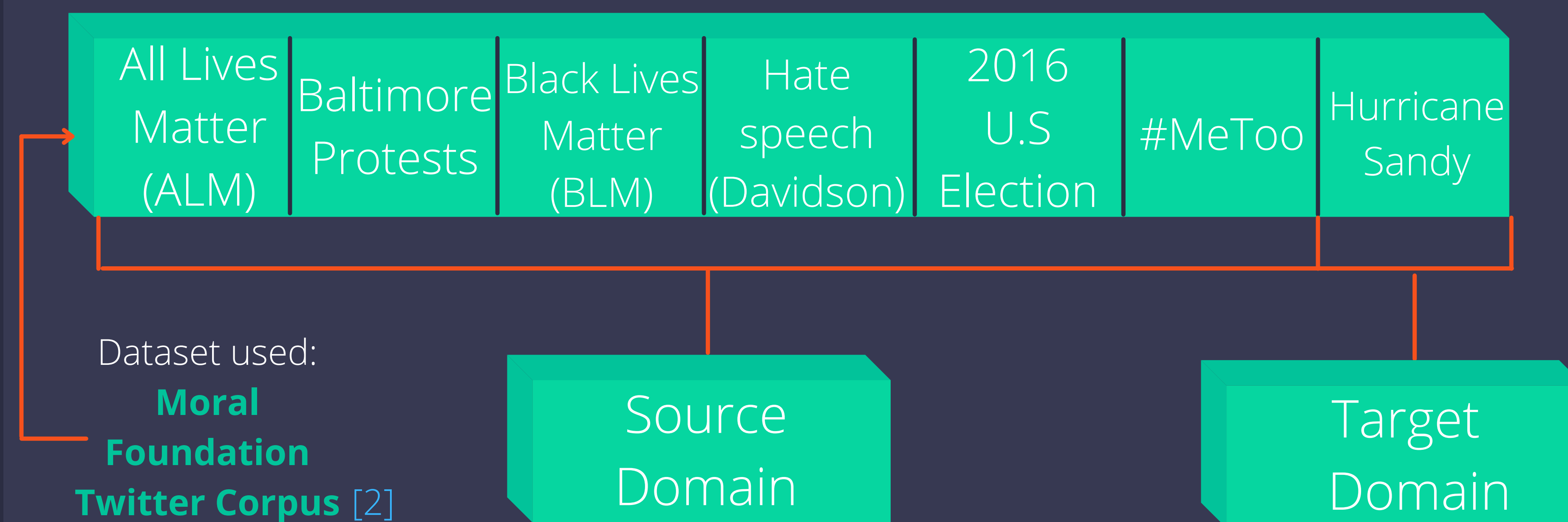


I. Background

- **Moral values** are the abstract motivations that define what is right or wrong
- Moral values can be expressed in terms of five **moral foundations** [1]
- A natural environment for analyzing morality is social media by employing **Natural Language Processing** (NLP) techniques
- **Research gap:** No study approaches the problem of **transferability** across domains of the studied models
- **Motivation:** Transferable models require **less data, less resources and usually lead to better performance**

II. Method



Evaluating transferability: Train a model on the source domain, use it to label the target domain and report performance through the **Macro F1-score**

- Each of the 7 sub-datasets will be chosen as the **target domain**, with the other 6 making up the **source domain**

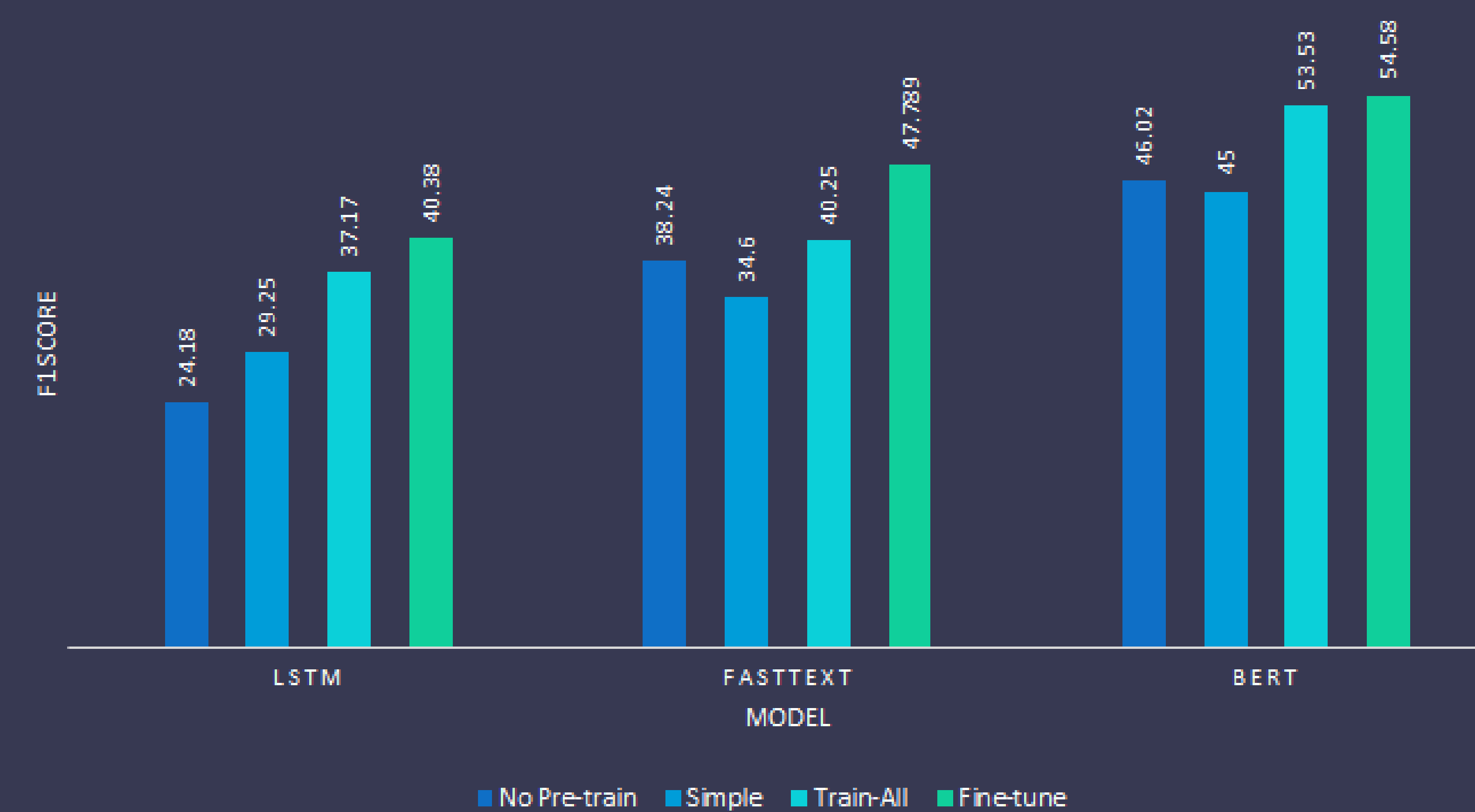
Three evaluated models: LSTM [3], FastText [4] and BERT [5]

Four different experiments:

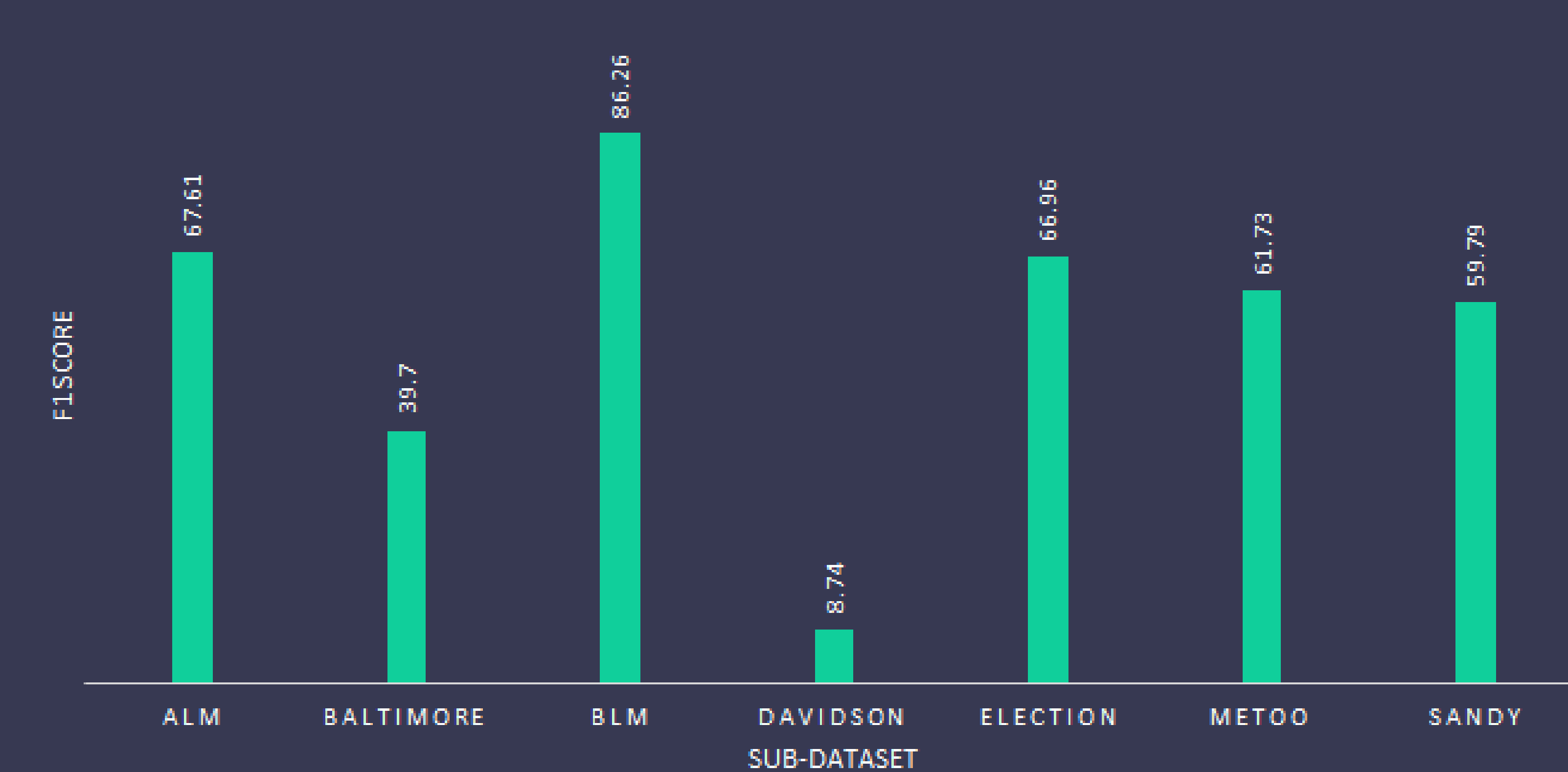
- **No pre-train** - Only train on the target domain
- **Simple** - Only train on the source domain
- **Train-all** - Train on a combination of the source and target domains
- **Fine-tune** - Firstly train on the source domain, then train on target domain

III. Experiments and results

MODEL PERFORMANCE COMPARISON



BERT PERFORMANCE PER SUB-DATASET



IV. Conclusions

- All models obtain a significant increase in performance by pre-training
- BERT outperforms both FastText and LSTM in all experiments
- BERT can generalize well to target domains
- Some domains are easier to predict than others

Contact: a.e.dondera@student.tudelft.nl

Acknowledgements: Special thanks to my supervisors for their support and guidance and to my colleagues for their continuous effort and help

V. References

- [1] J. Haidt and J. Graham, "When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize," *Social Justice Research*, vol. 20, no. 1, pp. 98–116, 2007
- [2] J. Hoover, G. Portillo-Wightman, L. Yeh, S. Havaladar, A. M. Davani, Y. Lin, B. Kennedy, M. Atari, Z. Kamel, M. Mendlen, et al., "Moral foundations twitter corpus: A collection of 35k tweets annotated for moral sentiment," *Social Psychological and Personality Science*, vol. 11, no. 8, pp. 1057–1071, 2020
- [3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997
- [4] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," *arXiv preprint arXiv:1607.01759*, 2016
- [5] J. Hoover, G. Portillo-Wightman, L. Yeh, S. Havaladar, A. M. Davani, Y. Lin, B. Kennedy, M. Atari, Z. Kamel, M. Mendlen, et al., "Moral foundations twitter corpus: A collection of 35k tweets annotated for moral sentiment," *Social Psychological and Personality Science*, vol. 11, no. 8, pp. 1057–1071, 2020