Exploring the Impact of Unlabeled Data on Job Salary Prediction

Anonymous

1 Introduction

This receive thousands of applications, among which fewer than 10 % have the appropriate skills. (Bhola et al., 2020)

thats why we always

2 Literature review

3 methods

4 Results

5 Discussion / Critical Analysis

6 Conclusions

References

Bhola, A., Halder, K., Prasad, A., and Kan, M.-Y. (2020). Retrieving skills from job descriptions: A language model based extreme multi-label classification framework. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5832–5842, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

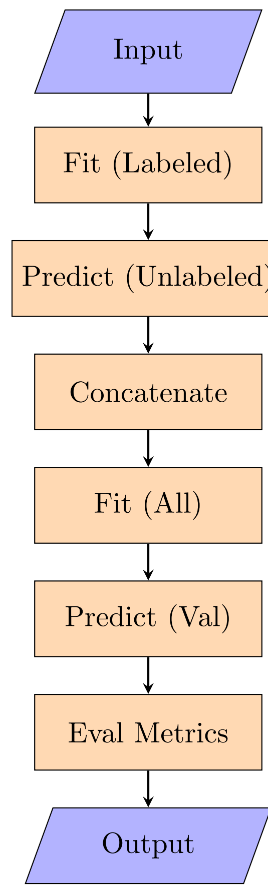

Figure 1. caption position

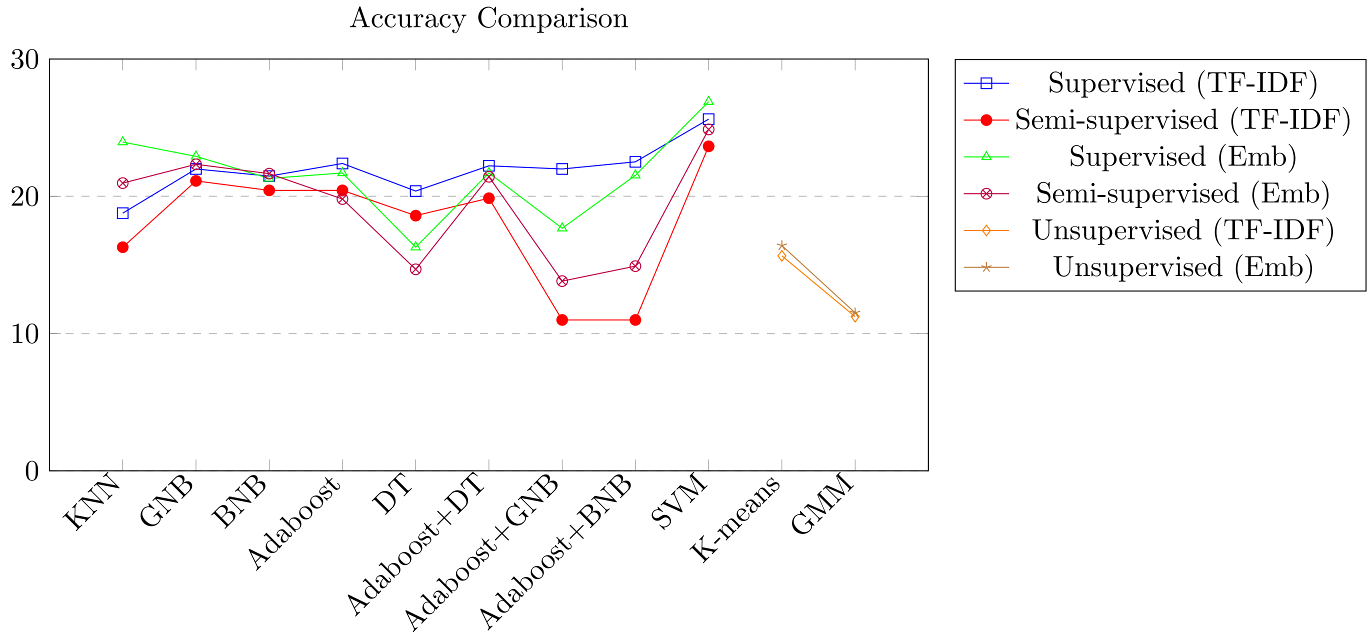

Table 1: Accuracy Comparison

Model	TFIDF		Embedding	
	Labaled data	Unlabeled data	Labaled data	Unlabaled data
KNN	18.77	16.29	23.95	20.96
GNB	21.99	21.12	22.91	22.33
BNB	21.47	20.43	21.3	21.65
Adaboost	22.39	20.43	21.7	19.8
Decision Tree (DT)	20.38	18.59	16.29	14.68
Adaboost + DT	22.22	19.86	21.7	21.42
Adaboost + GNB	21.99	10.99	17.67	13.82
Adaboost + BNB	22.51	10.99	21.53	14.91
SVM	**25.62***	**24.64***	**26.89***	**24.87***
K-means	-	**15.66***	-	**16.41***
GMM	-	11.23	-	11.51