PARSING & CLASSIFYING JOB POSTINGS

PURDUE UNIVERSITY | The Data Mine

AAHIL HUSSAIN, SUYASH JOHARI, BEN MOORMAN, MATT PAK

JOBVITE

INTRODUCTION

- Jobvite: recruiting software company that uses an applicant tracking software (ATS) to help thousands of companies' source, hire, and onboard top talent
- Motivation: Build separate features that allow for use to predict and improve the time to hire for different jobs.
- Goal: Create a model that can classify and parse different sections of a job posting into chunks based on its content.

TOOLS USED

- Python for Natural Language Processing (NLP), Data Analysis, and Modeling
- Microsoft Excel for Data Cleaning
- HuggingFace for Machine Learning
- Kaggle and Jobvite Datasets
- Brown (HPC Cluster)
- Git/GitHub for collaboration
- Jupyter Notebook for coding and computation

METHODOLOGY

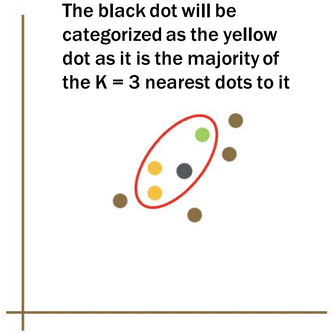
1. Split posting into chunks
2. Classify these chunks with categories to the right
3. Embed chunks using BERT
4. Train models below

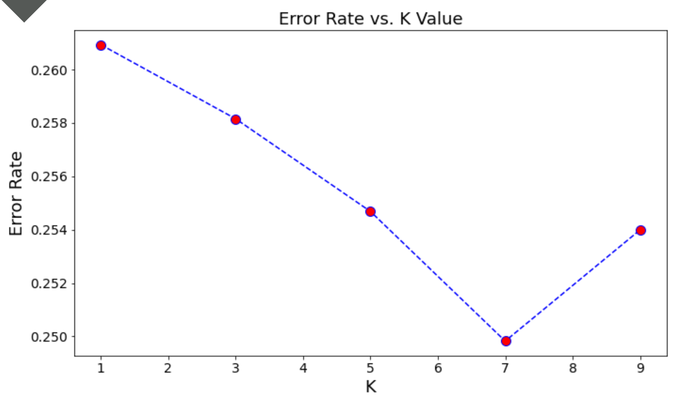
Models tested:

- K-Nearest Neighbors
- Random Forest Classifier
- Naïve Bayes

Parts of job posting:

- Company description
- Duties
- Skills
- Industry experience
- Education
- Diversity statement
- Pay
- Job type
- Location
- Header
- Benefits
- Bonuses





DATA

Kaggle Dataset (Indeed Job Postings)

- Job postings and related metadata
- About 30,000 records
- Split into chunks from each posting, by paragraph, lists, and new line spacings
- Labeled data with type listed in the Methodology section

[.3649]

Parse job posting using Python into "chunks"

Manually classify "chunks" of a job posting

Embed "chunks" using BERT

Train model to classify "chunks"

Model	Accuracy
KNN (k=7)	76%
Naïve Bayes	67.2%
Random Forest	53.1%

CONCLUSION + FUTURE GOALS

Deliverable:

- A trained NLP model that allows us to parse job postings
- Bert classification
  - We manually classified thousands of individual chunks
- Trained a KNN model

Next Steps:

- Create more training data
  - Using more training-data we can continue to update and refine the KNN model
- Implement these to continuously reduce time-to-hire for clients and future employees

REFERENCES

Thank you to Dr. Sasan Hashemi and Dr. Morgan Llewellyn for their mentorship these past two semesters.

- Link to Kaggle Data (https://www.kaggle.com/promptcloud/indeed-job-posting-dataset)
- Link to HuggingFace (https://huggingface.co)