REPORT FOR THE ASSIGNMENT

Name: Caleb pallapu **Email:** pallapucaleb@gmail.com **Phone:** 9908817990

Developed a comprehensive resume-job matching system that implements and compares two distinct machine learning approaches. The first model uses traditional TF-IDF vectorization combined with Random Forest classification, while the second leverages a modern BERT Mini transformer architecture with an additional neural network layer. After training both models on a resume dataset with match scores from 1-5, performed extensive evaluation revealing that the BERT-based model significantly outperformed the traditional approach, achieving 94.4% accuracy compared to 53.85%. The system features a Flask web application that allows users to input job descriptions and resumes either as text or file uploads, providing real-time match predictions from both models with confidence scores. Comparative analysis demonstrates the superiority of transformer-based models for semantic understanding in resume-job matching tasks, despite their higher computational requirements, ultimately delivering a production-ready solution for automated recruitment screening. We have also implemented a very basic

Models used

1. **TF-IDF + Random Forest**
2. **prajjwal1/bert-mini + Basic Neural Network**

Executive Summary

- BERT Mini + Neural Network: 94.4% accuracy (Excellent performance)
- TF-IDF + Random Forest: 53.85% accuracy (Moderate performance)
- Performance Gap: BERT is 40.55% more accurate than TF-IDF
- Recommendation: Use BERT Mini + Neural Network for production

Find models here: models

Classification Metrics

TF-IDF + Random Forest:

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\USER\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
Accuracy: 0.5385
              precision    recall  f1-score   support

           1       1.00      0.01      0.02        85
           2       0.53      0.53      0.53       372
           3       0.43      0.37      0.40       448
           4       0.49      0.66      0.56       622
           5       0.75      0.64      0.69       473

    accuracy                           0.54      2000
   macro avg       0.64      0.44      0.44      2000
weighted avg       0.57      0.54      0.53      2000

Model and vectorizer saved successfully!
```

prajjwal1/bert-mini + Basic Neural Network:

```
============================================================
🎯 FINAL ENHANCED BERT + NEURAL NETWORK RESULTS
============================================================
🏆 Best Validation Accuracy: 0.9373
📊 Final Test Accuracy: 0.9440

📈 Classification Report:
              precision    recall  f1-score   support

     Score 1       0.82      0.95      0.88        66
     Score 2       0.97      0.92      0.94       284
     Score 3       0.91      0.96      0.93       340
     Score 4       0.97      0.91      0.94       454
     Score 5       0.95      1.00      0.97       356

    accuracy                           0.94      1500
   macro avg       0.92      0.95      0.93      1500
weighted avg       0.95      0.94      0.94      1500


📊 Per-class Accuracy:
  Score 1: 0.955 (66 samples)
  Score 2: 0.915 (284 samples)
  Score 3: 0.959 (340 samples)
  Score 4: 0.907 (454 samples)
  Score 5: 0.997 (356 samples)

✅ Enhanced model trained successfully with prajjwal1/bert-mini
🔧 Architecture: BERT + Neural Network
📐 Hidden size: 256
```

Technical Insights

Why BERT Performs Better:

- Contextual Understanding: BERT understands semantic relationships
- Transfer Learning: Pre-trained on vast text data
- Neural Architecture: Can capture complex patterns
- Embedding Quality: Better representation of technical terms

TF-IDF Limitations:

- Bag-of-words approach: Loses word order and context
- No semantic understanding: Can't handle synonyms/related terms
- Feature sparsity: Limited to vocabulary seen in training

Conclusion

The BERT Mini + Neural Network model significantly outperforms the TF-IDF + Random Forest approach, demonstrating the power of modern transformer architectures for semantic text matching tasks. The 40.55% accuracy improvement justifies the additional computational requirements.

Final Decision: Deployed BERT Mini + Neural Network in production for all resume-job matching applications. The massive accuracy improvement provides substantial business value that outweighs the increased computational costs.