AI Project
22k-4233 Daniyal Saeed Dani
22k-4255 Ilyas Moiz
22k-4337 Muhammad Huzaifa

Resume Scanning & Interview

1. Project Motivation

- **Problem to Solve:**
  Traditional hiring processes are slow, subjective, and often inefficient. Recruiters manually scan resumes and conduct interviews, which can lead to biased or inconsistent decisions. We aim to automate resume screening and preliminary interviewing to assist recruiters by providing a standardized, data-driven initial hiring recommendation.

- **Strategic Goal Linked:**
  This project aligns with the strategic goal of improving efficiency, fairness, and accuracy in candidate evaluation, especially in preliminary rounds of hiring. It seeks to reduce recruiter workload and ensure candidates are evaluated consistently based on skill relevance.

---

2. Problem Definition

- **Specific Output to Predict:**
  The primary output is a **binary prediction**: *Hire* or *Do Not Hire*, based on resume-job matching and interview questionnaire scoring.

- **Input Data for the Algorithm:**

  - Resume (pdf)

  - Job Description (text file or parsed structured data)

  - A question bank (dataset of categorized interview questions from kaggle)

  - Candidate's answers to generated questions

- **Most Relevant Factors for Prediction:**

  - Match percentage between resume skills/experience and job description requirements

  - Quality and correctness of candidate answers to technical and behavioral questions

  - Decision tree thresholds for evaluating match strength and answer performance

---

3. Relevant Method/Model

- **Specific Output to Predict:**
  Again, predicting *Hire* or *Do Not Hire* based on combined matching score and interview performance.

- **Input Data for the Algorithm:**
  Same as above:

  - Resume text

  - Job description text

  - Data for decision tree to use for prediction / conclusion

- **Methods/Models Used:**

  - **Natural Language Processing (NLP)** techniques for resume-job matching (e.g., cosine similarity, keyword extraction)

- **Decision Trees** to:

    - Select relevant interview questions based on matched topics

    - Evaluate candidate answers by assigning weighted scores

    - Make final prediction

---

4. Performance Measurement

- **Measuring Accuracy of Predictions:**

    - Classification Accuracy: (Correct predictions / Total predictions)

    - Confusion Matrix: To analyze true positives (correctly hired), true negatives, false positives, and false negatives.

- **Minimum Expected Accuracy:**
  Since hiring is a subjective process and datasets might be limited, we expect at least **80% accuracy** for the decision tree model based on test data.

---

5. Risks and Dependencies

- **Risks:**

    - **Inaccurate Resume Parsing:** Poor parsing could lead to bad feature extraction.
      → **Mitigation:** Use robust NLP libraries (e.g., spaCy, transformers).

    - **Bias in Dataset:** If the question dataset is biased, the hiring decision may also be biased.
      → **Mitigation:** Carefully curate a diverse, inclusive dataset of questions.

    - **Overfitting the Decision Tree:** Overly specific decision trees may not generalize well.

→ **Mitigation:** Prune the tree and cross-validate.

- **Constraints:**

  - Limited dataset for real resumes, job descriptions, and answers.

  - Limited project time (semester project) to build and test models fully.

  - Need for ethical consideration: Avoid making final hiring decisions without human verification.

---

6. Run Performance Checks

- **Classification Accuracy:**

  - Calculate the number of correct *Hire/Do Not Hire* predictions divided by the total number of candidates processed.

Outline for Full Report (IEEE Format)

Abstract

- Problem: Inefficiency and subjectivity in early hiring stages.

- Solution: Automate resume matching and candidate evaluation using decision trees.

- Key Results: Target 80%+ prediction accuracy with confusion matrix analysis.

Introduction

- Describe the challenge in manual resume screening and interviews.

- Importance of automation in recruiting.

Background or Literature Review

- Review of resume screening techniques (NLP, keyword matching).

- Review of interview automation approaches (chatbots, question generators).

- Basics of Decision Trees in classification problems.

Methods and Materials

- Description of system components: Resume Scanner, Question Generator, Interview Evaluator.

- Explanation of how resumes and job descriptions are parsed and matched.

- How decision trees are constructed and pruned.

Data and Results

- Sample datasets (simulated resumes, job descriptions, questions).

- Training the model and evaluation results.

- Confusion Matrix and accuracy results.

Conclusion

- **Summary:** Our model successfully automates early hiring stages with reasonable accuracy.

- **Limitations:** Small dataset size, need for further fine-tuning for real-world usage.

- **Future Work:** Expand dataset, integrate deeper NLP models (like BERT) for matching, fairness auditing.