

Name: Ahmad Moneer Mzayek

Course Title: Graduation project

Course Code: Thesis_AI900

Lecturer: Dr.Ashraf Elsayed

Hand-in Method: Online Submission via E-Learning.

University: Midocean

Date: 16/4/2025

Assignment Report: Intelligent Job Recommendation System for Job Seekers Using Machine learning.

Graduation Project

Title: Intelligent Job Recommendation System Using Machine Learning

✓ 1. Dataset Description

To train and evaluate the proposed job recommendation system, three datasets will be used:

◆ A. Wuzzuf Job Listings Dataset (Egypt) – January 2025

- **Source:** Kaggle.
- **Link:** [Wuzzuf Job Dataset](#).
- **Type:** Real-world job listings from Egypt.
- **Size:** Over 8,000 job postings.
- **Features:** Job title, company, location, job type, career level, skills, job description.
- **Use in project:** Core dataset for analysing and recommending jobs.

◆ B. Resume Dataset – Gaurav Dutta

- **Source:** Kaggle
- **Link:** [Resume Dataset](#)
- **Type:** Classified resumes in plain text format
- **Size:** Several hundred resumes labeled by job domain (e.g., HR, Data Science, DevOps)
- **Use in project:** To build user profiles and train a recommendation model that matches resumes to job roles.

◆ C. LinkedIn Job Postings (2023–2024)

- **Source:** Kaggle
- **Link:** [LinkedIn Job Dataset](#)
- **Type:** Global job postings with detailed job descriptions
- **Use in project:** To enrich the job database and expand the recommendation scope.

✓ 2. Preliminary Literature Review and Next Steps

Recent research shows that hybrid recommendation systems—especially those combining content-based and collaborative filtering—perform better than single-method models. Studies such as:

- “Job Recommendation System using Content and Collaborative Filtering based Techniques” used TF-IDF and Word2Vec to improve job matching accuracy.
- “A Deep Learning Approach to Job Recommendation with NLP” showed the effectiveness of using text embedding and cosine similarity.

🔍 Next Steps:

- Pre-process resume and job data (cleaning, normalization).
- Use Word2Vec or TF-IDF to extract feature vectors.
- Build a content-based recommendation model that compares user profiles with job vectors.
- Evaluate the model using standard performance metrics.

✓ 3. Proposed Machine Learning Model

The proposed model is a **content-based filtering system that uses text vectorization (TF-IDF or Word2Vec)** to compare user profiles (resumes) with job descriptions.

🔧 How It Works:

1. Extract key skills from resumes and job descriptions.
2. Convert them into vector format using Word2Vec or TF-IDF.
3. Calculate **cosine similarity** between the user profile and job descriptions.
4. Recommend top-N matching jobs for each user.

This model is interpretable, lightweight, and suitable for a beginner-level implementation.

✓ 4. Timeline (8-Week Plan)

Week	Task
1	Dataset collection and environment setup.
2	Data cleaning and pre-processing.
3	Skill extraction and vectorization (TF-IDF / Word2Vec).
4	Model implementation (CBF with cosine similarity).
5	Model tuning and improvement.
6	Recommendation generation and validation.
7	Model evaluation (Precision, Recall, F1-score).
8	Report writing and final presentation preparation.

✓ 5. Evaluation Metrics

To evaluate the model's performance, the following metrics will be used:

- **Precision:** Percentage of recommended jobs that are truly relevant.
- **Recall:** Percentage of relevant jobs that were correctly recommended.
- **F1 Score:** Harmonic mean of precision and recall.
- **Cosine Similarity Score:** To measure similarity between resume and job vectors.

The model will be tested on a sample dataset and evaluated for accuracy and reliability.