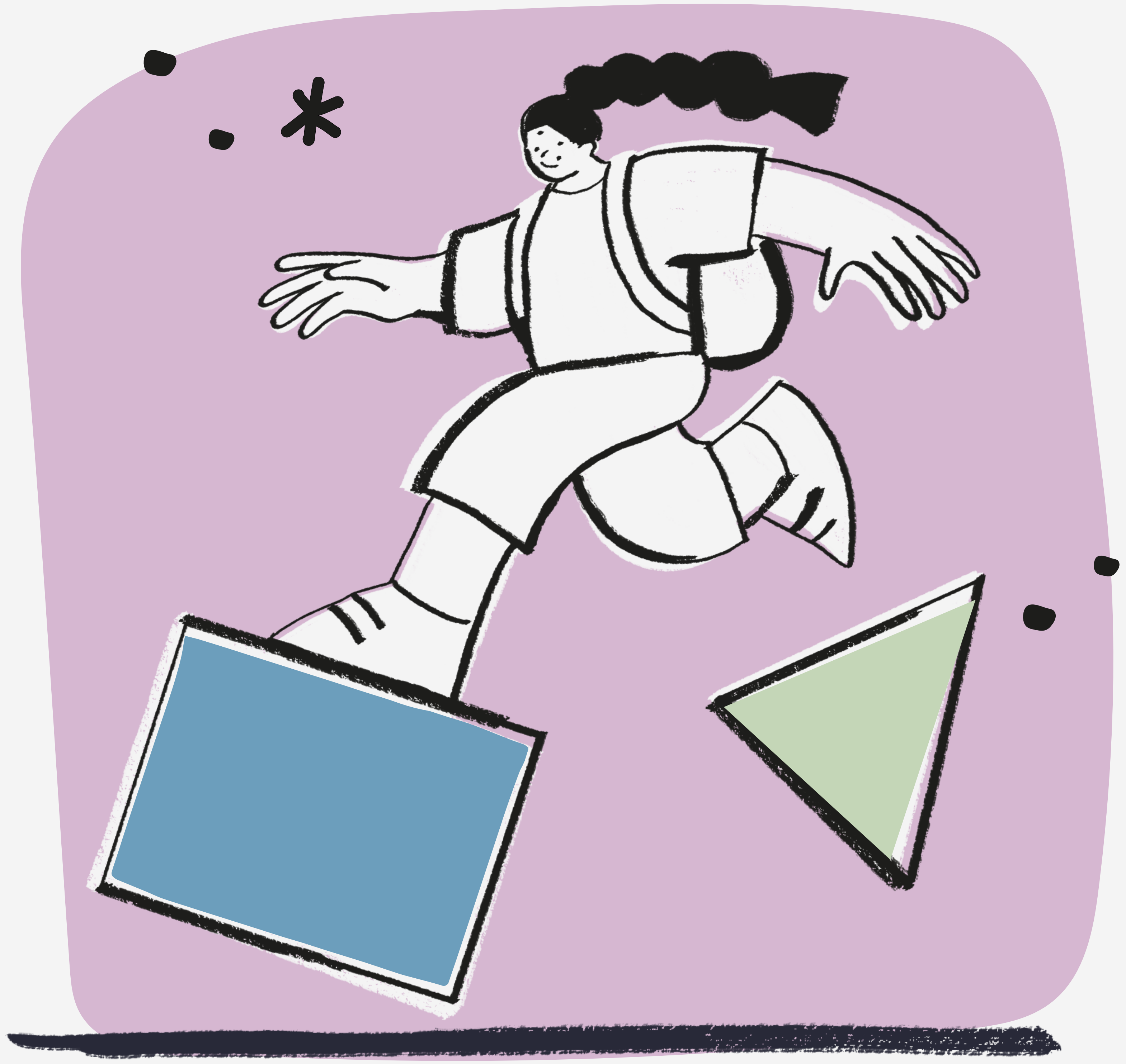
GRADE 5 · MATH

Advanced topics in machine learning

Lecturer:
Dr. Chen Hajaj
Team members:
Ziv Kalmer and Noa Nesher

The problem:

Navigating a saturated online job market, job seekers face the challenge of finding suitable positions that align with their skills and preferences amidst an overwhelming array of listings. Our project addresses the need for a streamlined and effective matching process to connect individuals with the right job opportunities.



Project Goals

- Improve Job Matching: Enhance the accuracy and relevance of job recommendations for seekers.
- Analyze Job Market: Understand patterns and trends within the employment landscape.
- Facilitate Employer Decisions: Provide tools for employers to efficiently find and attract suitable candidates.

Comparative Methods in Job Matching

A. Keyword Matching: Traditional method using simple keyword comparisons but lacks context sensitivity.

B. Resume Parsing: Extracts information from resumes for matching but may miss nuanced candidate attributes.

C. Collaborative Filtering: Recommends jobs based on similar user behavior but requires extensive user data.

D. Deep Learning Models: Utilize complex neural networks for pattern recognition but can be resource-intensive and require large datasets.

Dataset and Key Features

- Source: Careerjet, an extensive job search engine.
- Collection Method: Web crawling to bypass API limitations, ensuring comprehensive job description retrieval.

Features:

Job Titles

Company Names

Locations

Job Descriptions (detailed)

Required Skills

Experience Level

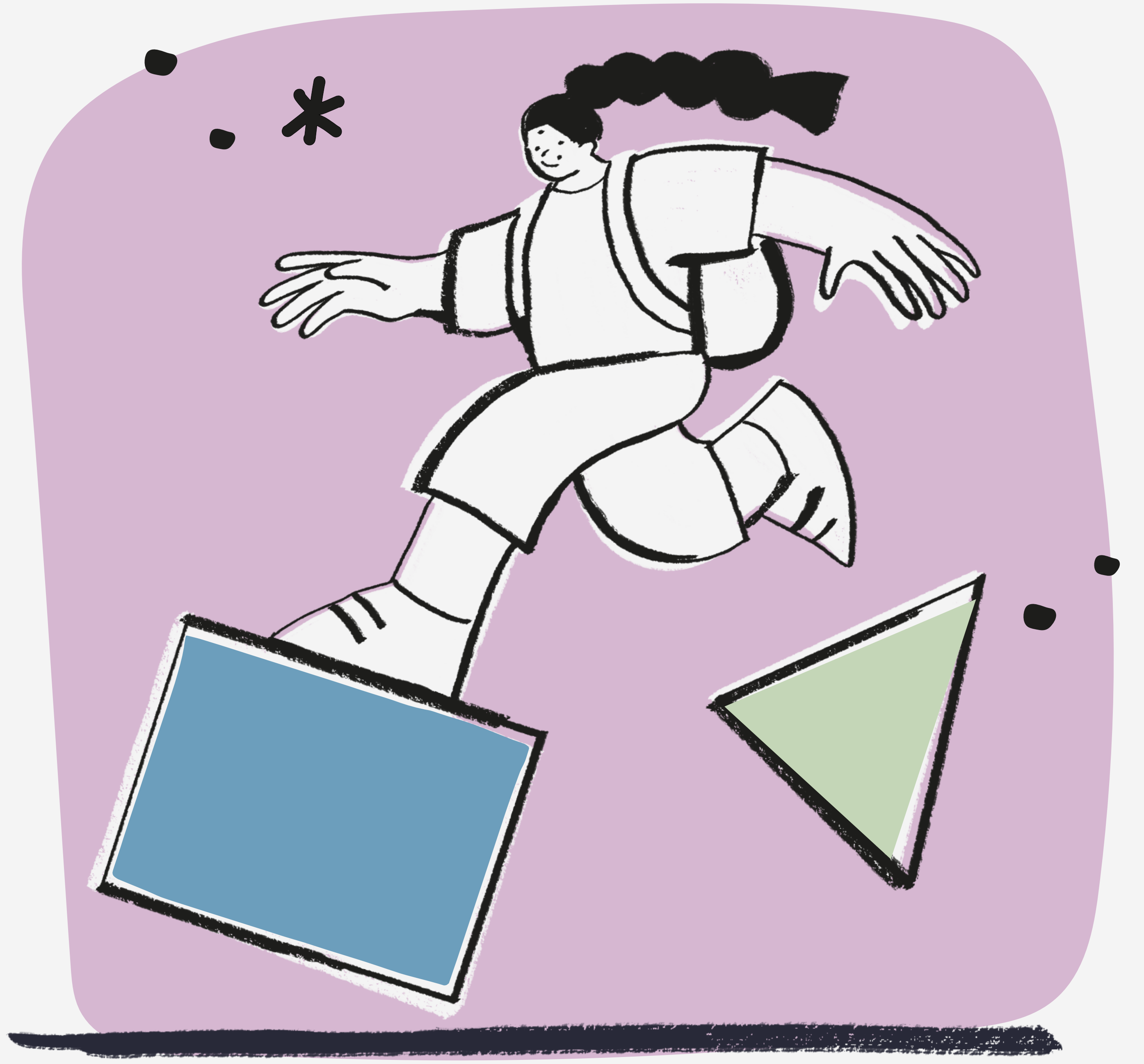
Methodology Overview

Algorithms and Techniques:

- Natural Language Processing (NLP): For parsing and understanding job descriptions and resumes.
- Word2Vec: To capture semantic relationships between words in job data, facilitating nuanced feature extraction.
- KMeans Clustering: Employed for segmenting job listings into distinct categories based on similarity.

Rationale:

- NLP and Word2Vec were chosen for their ability to deeply analyze and interpret textual data, crucial for accurate job matching.
- KMeans was selected for its efficiency and effectiveness in creating meaningful, manageable groups from complex datasets.



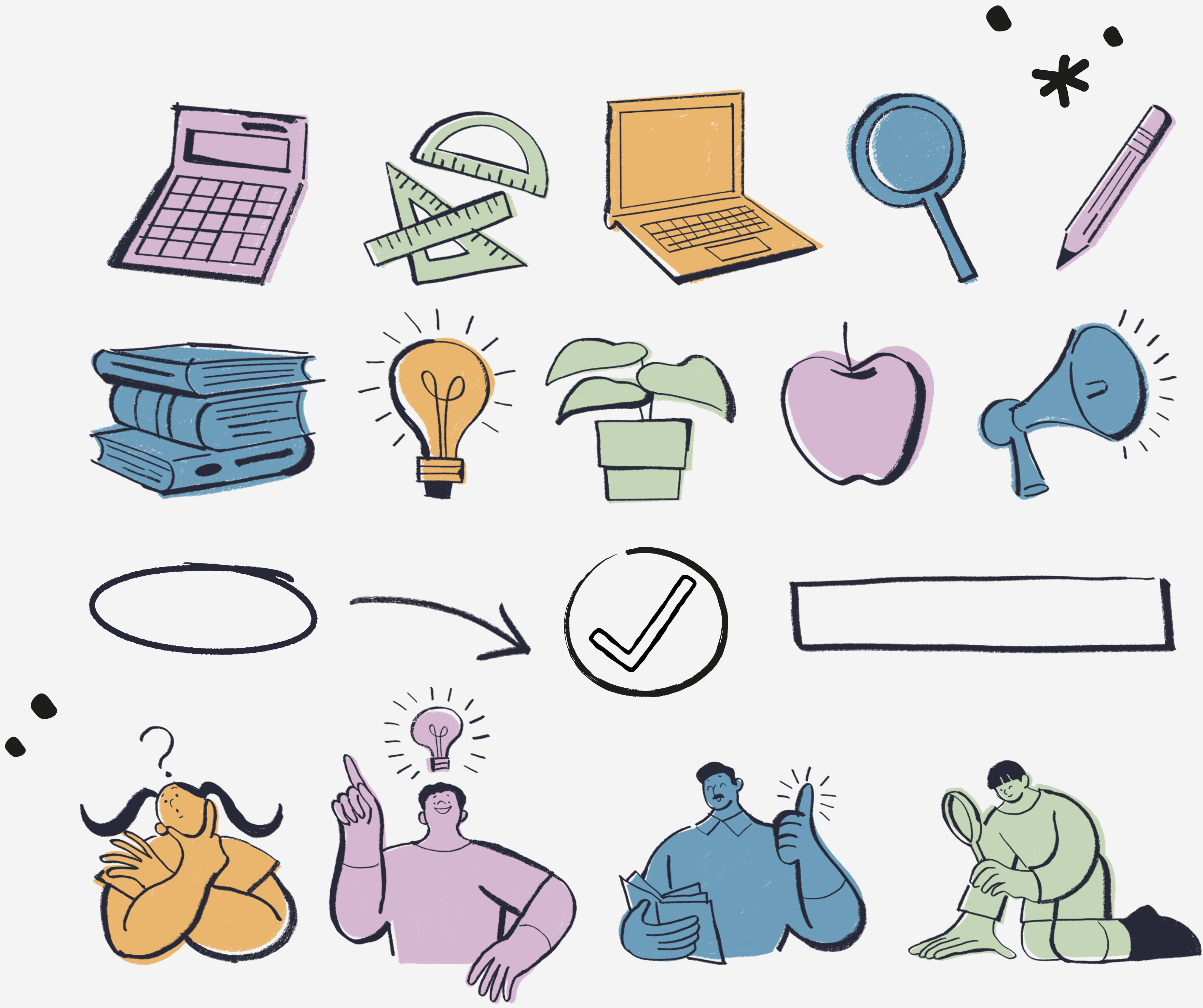
Experiments and Evaluation

<u>Experimentation:</u>

- Utilized KMeans to cluster job listings into distinct groups.
- Applied Word2Vec for semantic analysis of job descriptions.

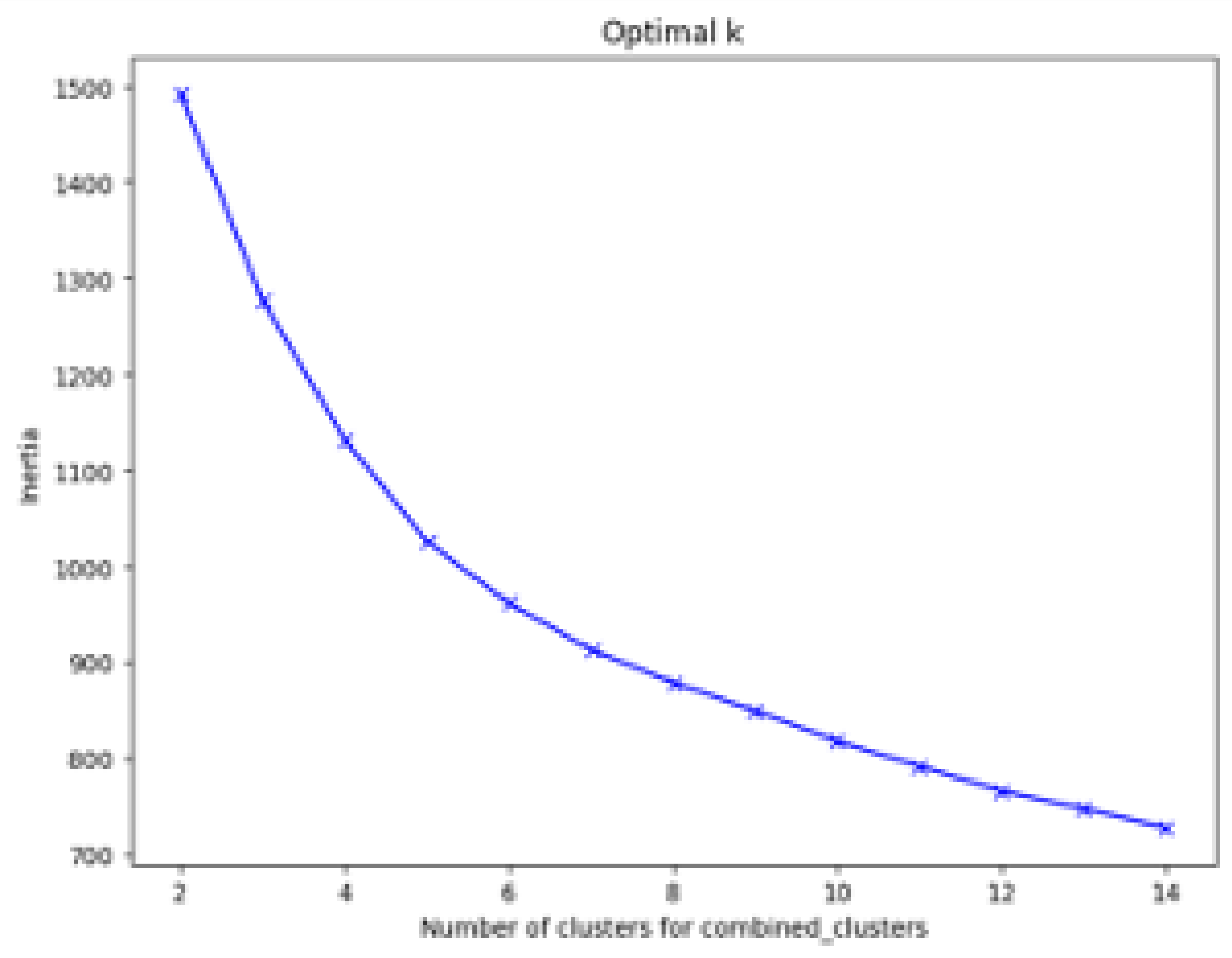
<u>Evaluation Techniques:</u>

- Elbow Method: Determined the optimal number of clusters for KMeans.
- Silhouette Score: Assessed the coherence and separation of clusters.
- t-SNE Visualization: Provided a visual representation of cluster distribution and separation.



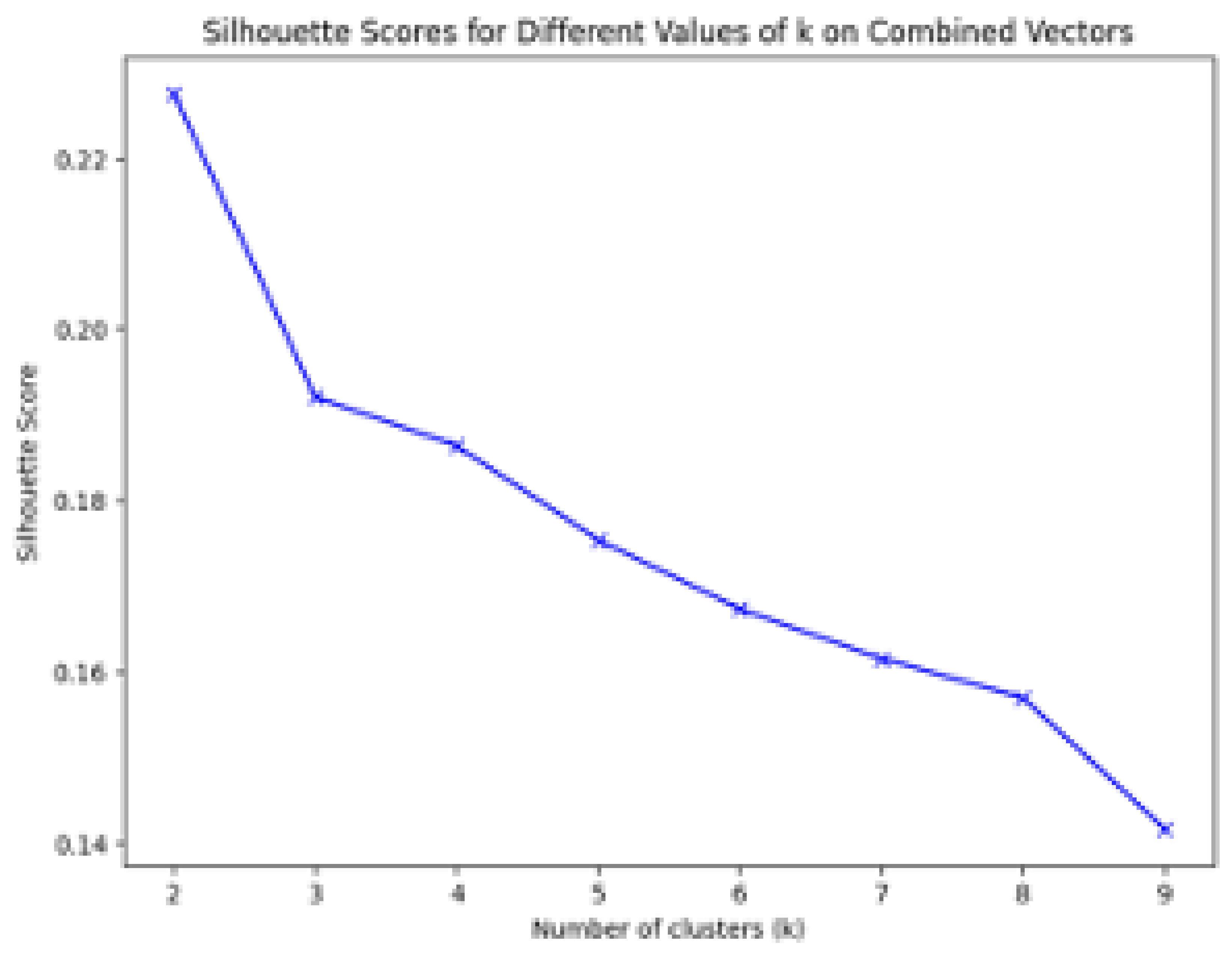
Results Overview

Elbow Method Plot



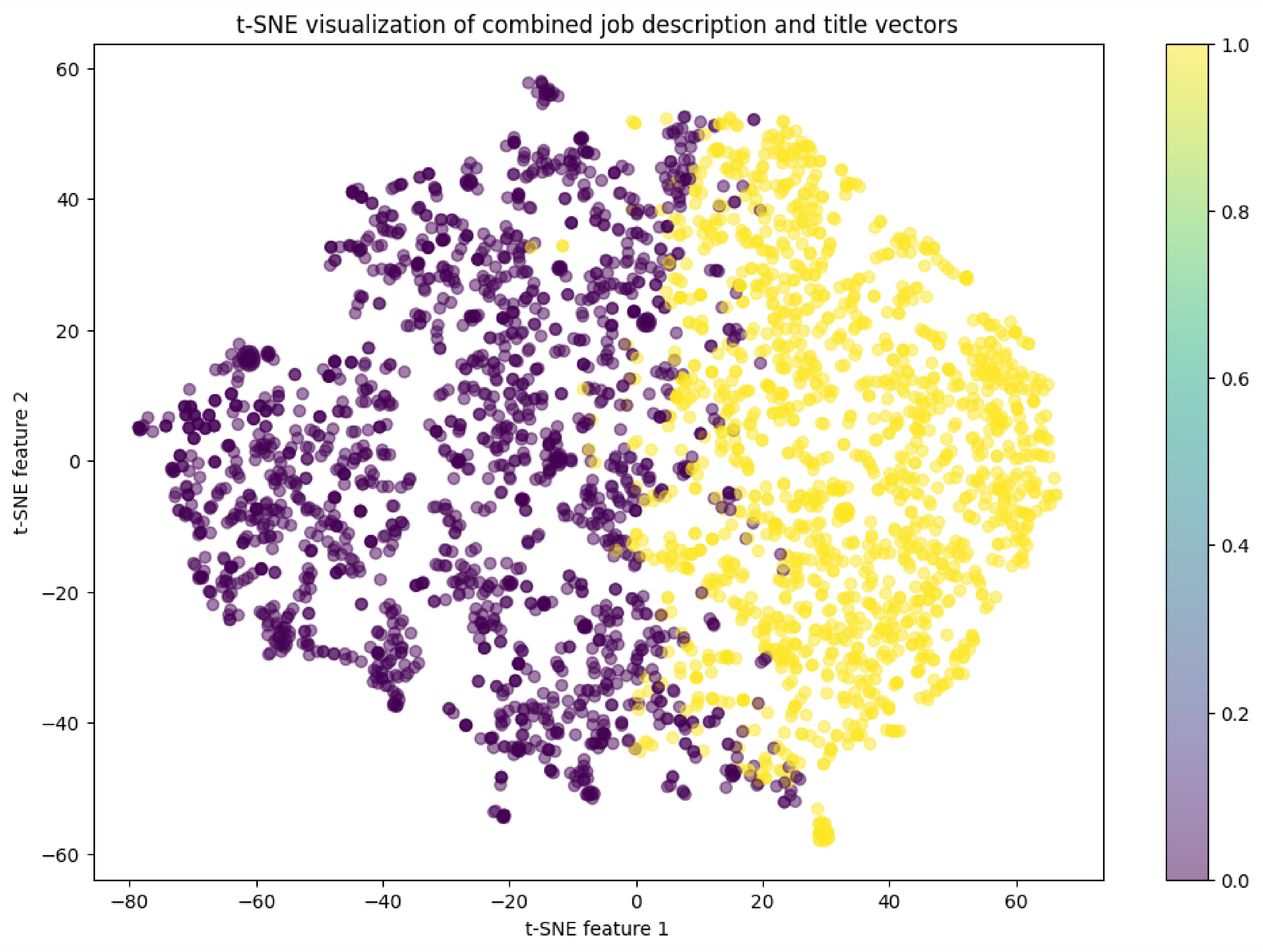
In the application of the KMeans algorithm, We utilized the Elbow Method to visually identify the optimal 'k'. The graph depicted a clear elbow at k=2, indicating a natural partition of the dataset. This finding was pivotal as it defined the granularity of our job market segmentation and directly impacted the specificity of our recommendations.

Silhouette Score Analysis



To evaluate the coherence of the clusters formed by KMeans, we relied on Silhouette Scores. This metric, ranging from -1 to +1, measures how close each point in one cluster is to points in the neighboring clusters. Our analysis yielded scores that were positive and closer to +1, especially pronounced at k=2, underscoring well-defined and distinct clusters. The positive scores across the board bolstered our confidence in the clusters' reliability and the model's utility in grouping similar jobs.

t-SNE Cluster Visualization



qualitative validation was provided by the t-SNE visualization, which allowed us to observe the distribution of clusters in a two-dimensional space. The visualization showed two distinct clusters with minimal overlap, corroborating our choice of k=2 from the Elbow Method and Silhouette Scores. The graph highlighted the clear demarcation between the clusters, reinforcing the efficacy of our feature selection and clustering process.

Technique Comparison

1. Word2Vec vs. Traditional Keyword Matching:
   - Word2Vec provides context-sensitive analysis, capturing semantic meanings, unlike keyword matching which only identifies exact word matches.

2. KMeans Clustering vs. Hierarchical Clustering:
   - KMeans is efficient for large datasets and clear cluster separation, whereas hierarchical clustering offers detailed dendrogram structures but is computationally intensive.

3. NLP Techniques (spaCy) vs. Manual Text Analysis:
   - Automated NLP with spaCy processes large text volumes quickly and accurately, outperforming time-consuming and less consistent manual analysis.

Conclusion

- We developed a job recommendation system using machine learning to categorize and match job listings with candidate profiles.
- We employed NLP and clustering algorithms to process and structure the job market data, validating our approach with solid evaluation metrics.
- Our project improved job search efficiency and offered insights into the job market's dynamics, laying groundwork for further advancements in this field.