



Resume Reveal

Reveal your seniority level

By:
Matan Cohen
Shira Shani
Eden Menahem



Description

“Resume Reveal”

Accurately assessing candidate seniority from resumes is a critical yet challenging task, complicated by the prevalence of overstated experience and ambiguous self presentation.

Task

- Input: Resume text + Job role.
- Output: Predicted seniority level -Junior / Mid / Senior.
- Task: Multi-class classification (single-label prediction).

Data and evaluation

- **Dataset:** 523 synthetic resumes and web-scraped resumes.
- **Labels:**
 - LM-generated resumes → Generated Based on Seniority and job-role
 - Web-Collected resumes → Include Job Role + Inferred Seniority

Prior Art

TITLE	<u>ResumeAtlas: Revisiting Resume Classification with Large-Scale Datasets and Large Language Models</u>	<u>Construction of English Resume Corpus and Test with Pre-trained Language Models</u>	<u>conSultantBERT: Fine-tuned Siamese Sentence-BERT for Matching Jobs and Job Seekers</u>
Approach / Model	Used BERT and compared it to TF-IDF for classifying resumes into Junior, Mid, and Senior levels.	Used resume text, splits it into parts like experience or education, and uses BERT or DistilBERT to predict the correct label for each part. Results were compared to TF-IDF.	Fine-tuned a BERT model to match resumes with job descriptions, and compared it to TF-IDF using similarity scores.
Data	13,389 resumes labeled by seniority and job role.	1,484 resumes: 286 labeled, 1,198 from OCR.	270,000 resume-job pairs from real applications.
Metrics	Checked how often the top 1 or top 5 predictions matched the correct seniority.	Measured how well each part of the resume (like experience) was labeled correctly.	Compared how well resumes matched jobs using accuracy and similarity scores.
Results	BERT got better accuracy than TF-IDF (92% vs. 85.8%).	DistilBERT worked best. Experience helped the model.	SBERT gave better job-resume matches than TF-IDF.

Steps

Preprocessing

A combination of 2 types of data:

- A synthetic dataset of resumes was generated iteratively using controlled guidelines. each resume was designed to reflect the unique job role, experience level (junior/mid/senior).
- resume samples were scraped from an online source - “<https://www.hireitpeople.com/>”

Labeling

- Labels :Junior, Mid and Senior.
- Seniority was inferred from context and years of experience in the relevant job role.

Modeling

- Baseline – Logistic Regression (TF-IDF) Trained on TF-IDF word-vector features simple, fully interpretable benchmark model used to quantify the added value of context-aware language understanding.
- BERT-base-uncased is a pretrained model fine-tuned for multiclass classification. It jointly encodes the job role and the resume text in a unified context. Captures deep semantic links across experience, responsibilities, and role-specific terms. Delivers accurate Junior/Mid/Senior predictions.

Evaluation

- Main metrics: Accuracy, F1-score.
- Additional metrics: Confusion matrix, loss monitoring (cross-entropy).

EDA

Exploration and Baseline

Resume Dataset Summary:

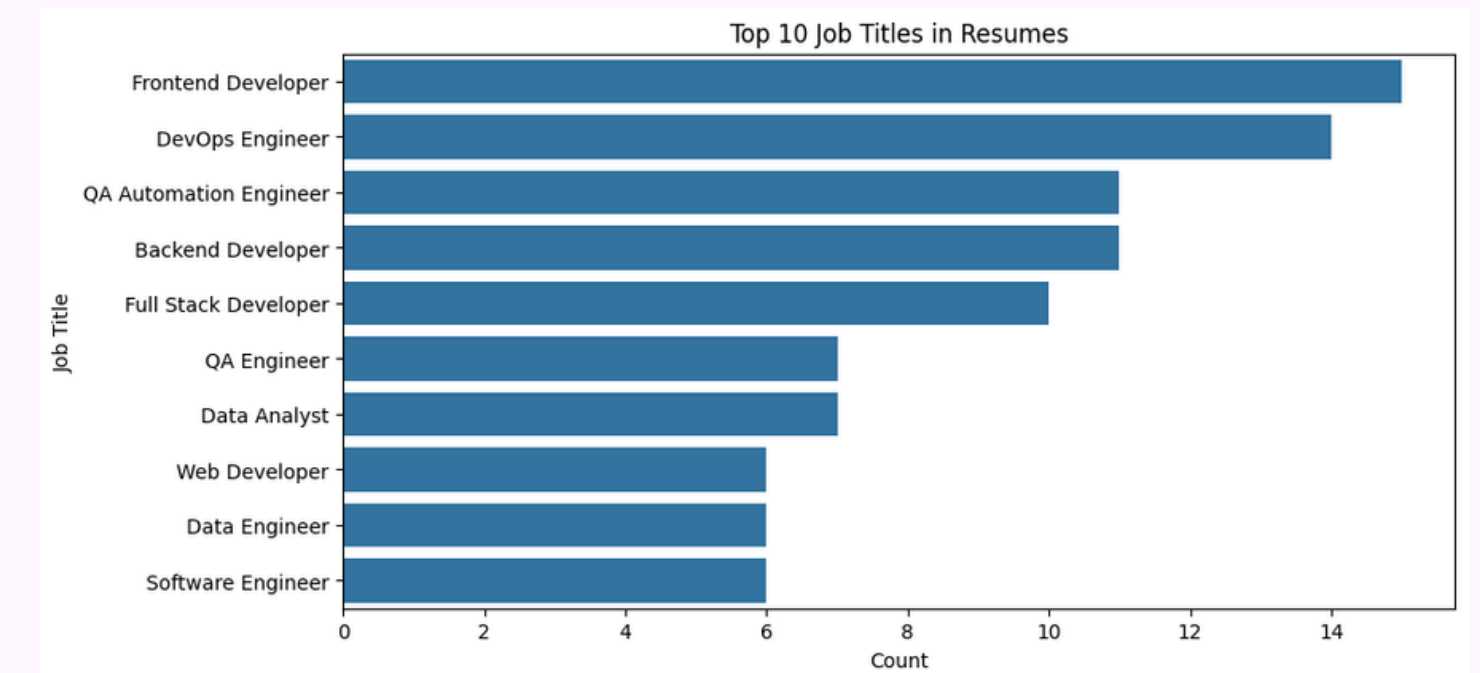
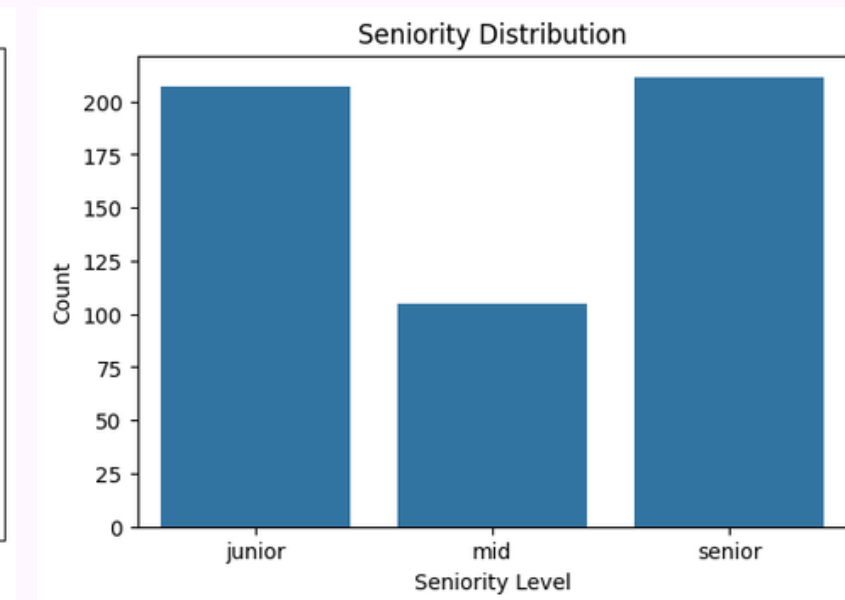
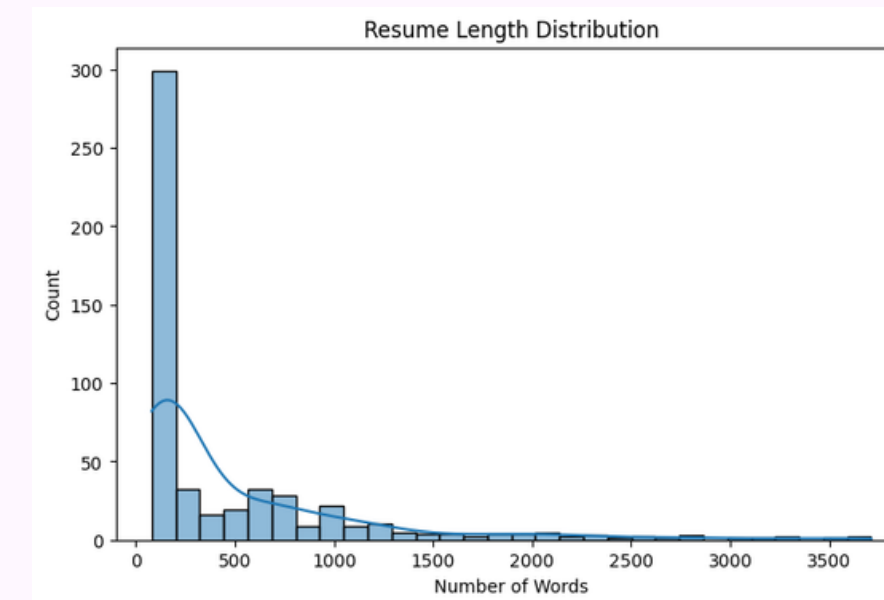
- Total resumes: 523
- Label distribution: {'senior': 211, 'junior': 207, 'mid': 105}

Resume Length Statistics:

- Mean: 495.9 words
- Std: 632.96 words
- Min: 80, Max: 3712
- Most resumes are between 133 and 663 words long (middle 50%)

Top 10 job roles:

- Frontend Developer: 15 resumes
- DevOps Engineer: 14 resumes
- QA Automation Engineer: 11 resumes
- Backend Developer: 11 resumes
- Full Stack Developer: 10 resumes
- QA Engineer: 7 resumes
- Data Analyst: 7 resumes
- Web Developer: 6 resumes
- Data Engineer: 6 resumes
- Software Engineer: 6 resumes



Exploration and Baseline

Baseline

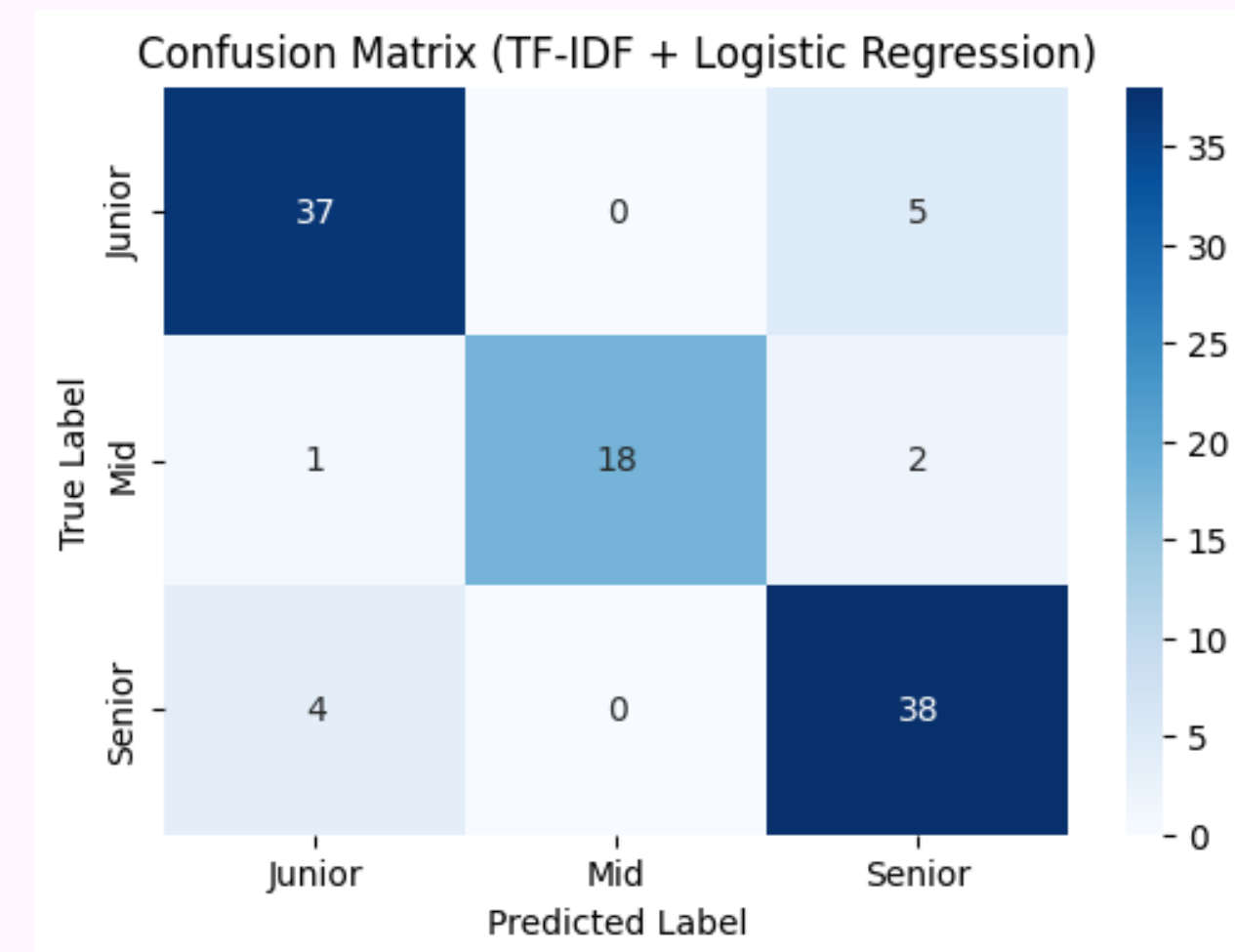
- Combined job role and resume text into a single input
- Applied TF-IDF vectorization (1–2 n-grams, max 5000 features)
- trained a Logistic Regression classifier
- Evaluation done on a stratified 80/20 train-test split by seniority
- Metrics: Accuracy and F1-score per class

Evaluation Results:

Accuracy: 88.57%

F1-score per class:

- Junior: 0.88
- Mid: 0.92
- Senior: 0.87



Insights/Recommendations

What we learn:

- The baseline model using TF-IDF and logistic regression reached 88.6% accuracy.
This suggests that even a basic approach can identify patterns related to seniority levels in resumes.
- BERT-based models from previous work have consistently outperformed traditional methods.
- Our EDA testing revealed significant imbalances in the dataset that could lead to overfitting.

Next steps/improvements:

- Add more resumes to the data.