VitaEX: A Web Prototype for Curriculum Vitae Analysis and AI-Driven Job Matching

Abdiel Reyes Vera[1,2], Oscar Huerta Villanueva[1], Diego Martínez Méndez[1](✉), and Luis Fernando Valle Hernández[1]

[1] Escuela Superior de Cómputo, IPN, Mexico City, Mexico
✉ dmartinezm1707@alumno.ipn.mx
[2] Centro de Investigación en Computación, IPN, Mexico City, Mexico

Abstract. Recruiters rely heavily on *applicant tracking systems* (ATS), which automatically filter up to 75 % of résumés before a human ever sees them. VitaEX is a three-tier, Django-based web prototype that helps undergraduate and recently-graduated students in Artificial Intelligence (AI) tailor their curricula vitae (CVs) to pass these filters while simultaneously recommending the most relevant job vacancies. This article synthesises every major facet of the original Spanish thesis: literature review, dataset construction via large-scale LinkedIn scraping, an NLP pipeline, an Apriori rule-mining recommender, spiral & CRISP-DM methodologies and architecture.

Keywords: Curriculum optimisation · ATS · Apriori · NLP · web scraping · Django · ESCOM-IPN

1 Introduction

Applicant tracking systems have become highly relevant: Fortune 500 companies deploy them in 98 % of hires [1], often eliminating competent "hidden workers" whose CVs lack the *exact* keywords a vacancy demands. Mexico's National Graduate Survey (2023) shows 46.3 % [2] of graduates deem job-hunting "difficult" — largely due to missing experience and ATS incompatibility. VitaEX attacks this gap with an AI-powered CV analyser plus a vacancy recommender specialised in AI roles across five Mexican tech hubs.

2 Related Work

Table 1 benchmarks eight commercial and academic systems. Most either (i) generate generic résumés, (ii) recommend jobs without explaining the match, or (iii) ignore student-specific constraints. ResumeNet [3] scores CV quality but not vacancy fit; GIRL [4] uses large language models but lacks transparency. In contrast, VitaEX combines interpretable Apriori rules with a student-centered workflow and provides a free usage tier.

Applications and Research Works	Recommends based on job offers	Generates CV templates	Uses Machine Learning	Job offer recommendation	Student-oriented and free
Application of NLU methods for CV recommendation [5]	✓	✗	✗	✗	✗
Generative Job Recommendations with LLM (GIRL) [6]	✓	✗	✓	✗	✗
ResumeNet [3]	✗	✗	✓	✗	✗

Neural networks for recruitment [7]	✗	✗	✓	✗	✗
LinkedIn Resume Builder [8]	✓	✓	✗	✗	✓
CVapp [9]	✗	✓	✗	✗	✗
Jobania [10]	✓	✓	✗	✗	✗
CVMATCHER [11]	✓	✓	✓	✓	✗
VitaEX (proposed system)	✓	✓	✓	✓	✓

Table 1: Comparison of systems and research works related to CV analysis and job recommendation

3 Project Overview

3.1 Objectives

a) Scrape and normalise ~6 000 AI vacancies (Nov 2024–May 2025).
b) Build an NLP module to extract skills, experience and job context from both CVs and postings.
c) Train an Apriori-based recommender to score CV–vacancy affinity and generate concrete rewriting tips.
d) Deliver a responsive Django/Bootstrap interface for students, companies and administrators.

3.2 Methodologies

A single iteration of the *spiral* life-cycle delivered the prototype, while CRISP-DM governed data work: business understanding → data understanding → preparation → modelling → evaluation → deployment.

4 Dataset Construction

4.1 Web Scraping Pipeline

Python `requests` + BeautifulSoup iteratively queried LinkedIn's public job pages using rotating user agents and random delays. Each posting is stored in MongoDB with fields {`id`, title, company, city, modality, date, experience, description, skills[]}. Monthly incremental runs refresh the corpus and trigger rule re-mining.

4.2 Pre-processing

Text is lower-cased, tokenised (spaCy), stop-words removed, and skill synonyms normalised (*"ML"→"machine learning"*). Eight AI-centric job titles (AI Engineer, NLP Engineer, . . .) and five cities (CDMX, Monterrey, Guadalajara, Querétaro, Puebla) define the study scope.

5 Recommender

5.1 Apriori Rule-Mining

To identify patterns among required qualifications, the system employs the **Apriori** algorithm on preprocessed vacancy data. Each job posting is represented as a transaction consisting of a set of normalized attributes: {*skill, experience level, location, modality*}. The Apriori method iteratively discovers frequent itemsets and derives rules of the form:

$$\text{Antecedents} \Rightarrow \text{Consequents}$$

where antecedents are conditions commonly appearing together (e.g., "Python", "CDMX"), and consequents are recommendations inferred from these patterns (e.g., "TensorFlow").

Rules are filtered using three metrics:

- **Support:** The proportion of job postings where the rule occurs.
- **Confidence:** The conditional probability of the consequent given the antecedent.
- **Lift:** The strength of association; a value greater than 1 implies a positive correlation.

To build a transparent recommendation engine, the system transforms each student profile into a transaction $\mathcal{T}$ containing elements such as `skills`, `experience_level`, `city`, and `modality`. Apriori association rules are generated with thresholds of **support** $> 1\%$, **confidence** $> 50\%$, and **lift** > 2.

Out of a total of **2,075,096** candidate rules, only **533,889** were retained after filtering. This included the removal of **1,541,207 redundant or uninformative rules**, such as those whose antecedents and consequents overlapped or involved unspecified attributes. Each surviving rule is interpretable as a student recommendation, such as:

```
If the student has "Data Analysis" and "Intermediate Experience" ⇒ Add "Machine Learning" to
improve fit (Confidence: 91%).
```

The system ranks job vacancies by combining cosine similarity (from SBERT embeddings) with rule confidence. Suggestions are generated for any missing but high-lift consequent features.

Algorithm Workflow:

1. The student profile is converted into a set of normalized attributes.
2. Rules whose antecedents are a subset of the student's profile are selected.
3. The system recommends any missing attributes from the consequents.
4. Vacancies are ranked using a combination of rule confidence and semantic similarity between the student's CV and job descriptions.

5.2   Results

The discovered rules reveal various insights. For instance, some rules with **100% confidence** show moderate lift, indicating common but not highly distinctive associations (e.g., "SQL" ⇒ "Python"). In contrast, other rules with **high lift** (above 3.0) suggest more meaningful and less obvious relationships (e.g., "Pandas", "Querétaro" ⇒ "TensorFlow").

- **Figure 1** presents the top 10 association rules sorted by lift, translated and aggregated for readability.
- **Figure 2** shows the most frequent attributes in the dataset, excluding vague or overly generic terms.
- **Figure 3** displays a variety of rules by confidence, ranging from 50% to 100%, offering both precise and broadly useful suggestions.

These results directly feed the recommender system. For example, if a student from CDMX lists "SQL" and "Pandas" but not "TensorFlow", the system recommends adding "TensorFlow" based on a strong rule with both high confidence and lift, thereby improving the candidate's alignment with prevalent industry patterns.
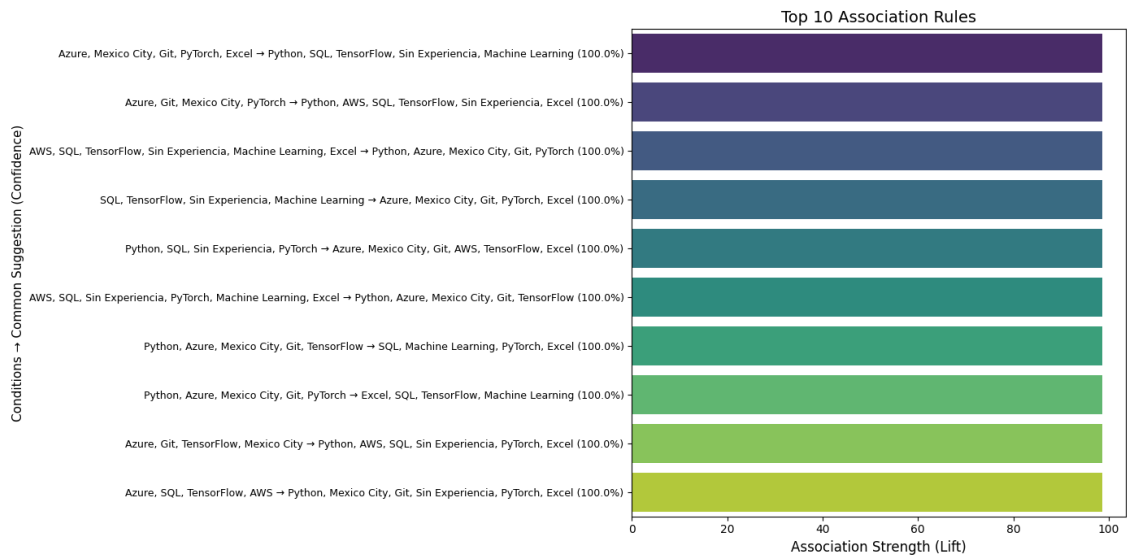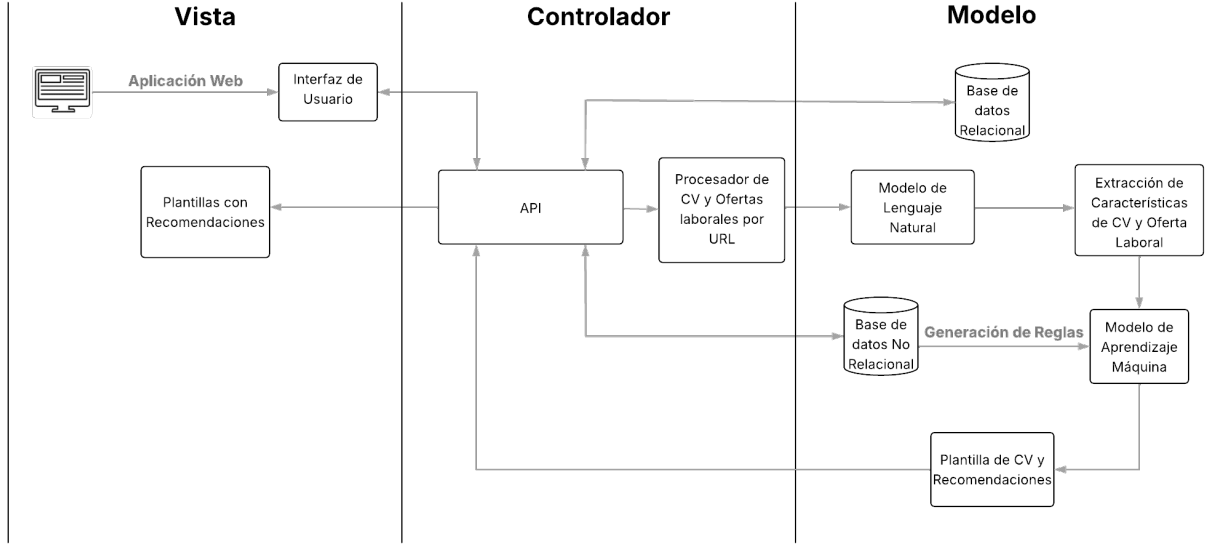

Fig. 1. Top 10 association rules by lift.
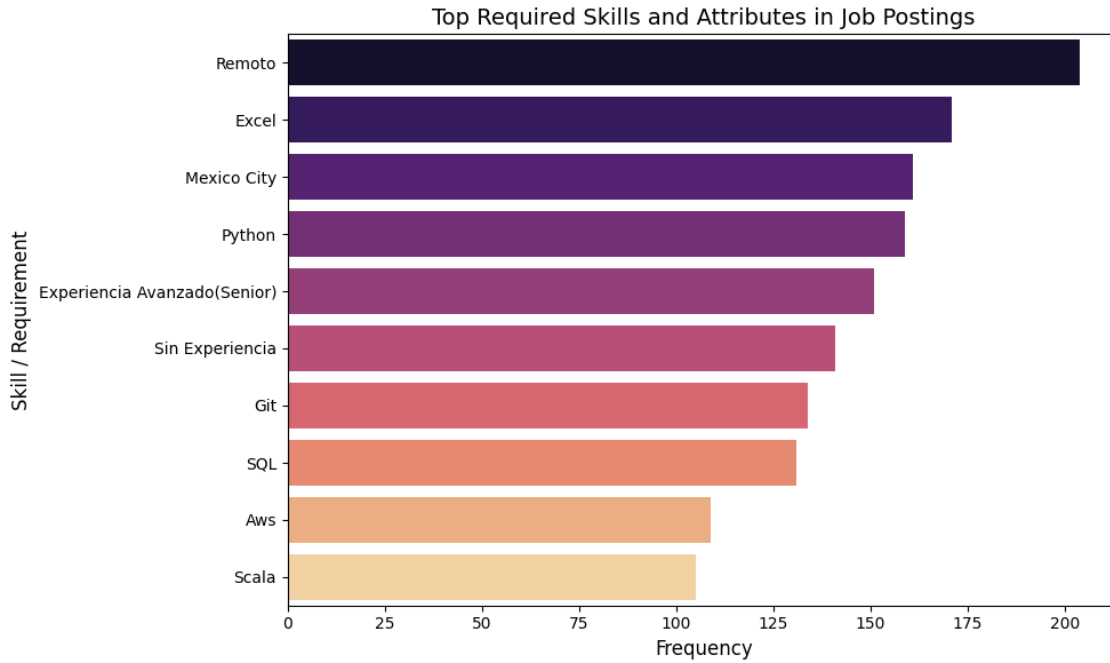
Fig. 2. Most frequent attributes detected in job postings.
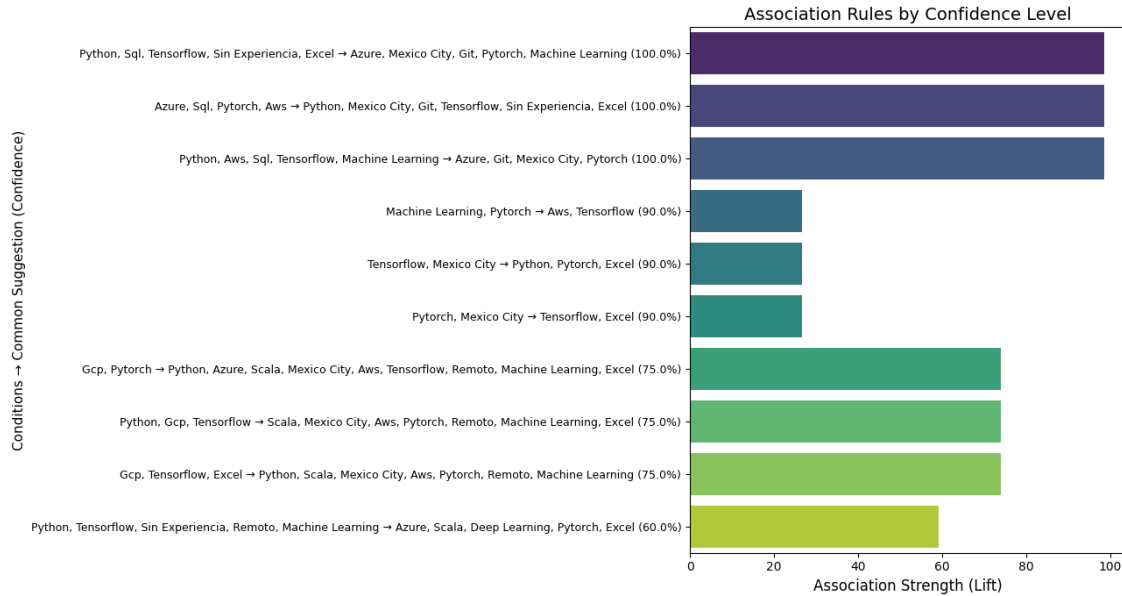
Fig. 3. Confidence levels of extracted rules.

6   System Architecture

Figure 4 shows a three-layer MVC stack: **View** (Bootstrap 5) → **Controller** (Django REST API) → **Model** (NLP services + MongoDB/PostgreSQL). Celery workers handle asynchronous scraping and model retraining.

Fig. 4. High-level architecture of VitaEX.

7   Conclusion

VitaEX demonstrates that combining transparent rule-based systems with modern NLP can materially improve employment prospects of AI students in Mexico. Our innovative application of the **Apriori algorithm** has proven particularly valuable for uncovering hidden skill associations in the job market and generating actionable CV recommendations. By mining over 500,000 association rules from real vacancy data, the system provides explainable suggestions that help candidates optimize their CVs for ATS compatibility while identifying relevant career opportunities.

Future work will focus on enhancing the **keyword extraction capabilities** for user-submitted CVs through:

1. Advanced context-aware NER models to better identify implicit skills
2. Extraction from CV formats (.docx or.pdf)
3. Skill normalization using domain-specific ontologies
4. Temporal analysis of skill importance

These enhancements will further bridge the gap between academic preparation and industry requirements in Mexico's rapidly evolving AI job market.

Acknowledgements

Bibliography

1. H. B. School and Accenture, "Hidden workers: Untapped talent," 2021.

2. U. del Valle de México (UVM), "Encuesta nacional de egresados 2023." UVM Opinión Pública, nov. 2023. [En línea]. Disponible en: https://opinionpublica.uvm.mx/wp-content/uploads/2023/11/BROCHURE_ENE-2023-1.pdf, 2023.

3. J. Hu *et al.*, "Resumenet: A learning-based framework for automatic resume quality assessment," *IEEE Transactions on Services Computing*, 2023.

4. A. Shen and B. He, "Generative job recommendations with large language models," *arXiv preprint arXiv:2401.01234*, 2024.

5. S. M. R. Cuevas, "Aplicación de métodos nlu en la recomendación de cvs para la selección de personal." Trabajo de Fin de Grado, Universidad de Valladolid, 2022.

6. Y. Yang, W. Li, and B. Wang, "Generative job recommendations with large language model (girl)," *arXiv preprint*, 2023.

7. G. S. Franco, M. A. G. Pérez, M. A. G. Silva, and V. M. Z. García, "Redes neuronales: nueva estrategia de inteligencia artificial para implementar dentro del proceso de reclutamiento y selección de personal." Universidad Politécnica Metropolitana de Hidalgo, 2018.

8. "Linkedin resume builder | linkedin help." https://www.linkedin.com/help/linkedin/answer/a551182. Accedido: 8 de marzo de 2025.

9. CVMATCHER, "Cvmatcher: Encuentra el trabajo perfecto para ti." https://www.cvmatcher.app/, 2025. Accedido: 8 de marzo de 2025.

10. "Crea tu currículum vitae gratis y encuentra trabajo en 2025." https://cvapp.mx/, 2025. Accedido: 8 de marzo de 2025.

11. Jobania, "Analizador de compatibilidad de currículum y oferta laboral." https://www.jobania.cl/analizar-cv-oferta, 2025. Accedido: 8 de marzo de 2025.