Project Report: Website for Algeria Venture Administration

1. Project Overview

The project involved developing a website for the administration of Algeria Venture, a government initiative supporting startups and entrepreneurship. The primary goal was to create a platform where administrators could upload a CSV file containing data about proposed jobs and projects. The system would then filter the data based on specific criteria to identify auto-entrepreneurial jobs that align with the ministry's requirements. After the AI filtering process, a committee (لجنة) would review the results and make the final decisions, ensuring a hybrid approach combining AI and human expertise.

Key Objectives:

1. **Filtering Data:** Automatically filter out jobs that require a "Registre de Commerce" or "Bita9a Fanya" (technical approval).
2. **Exclusion of Accepted Jobs:** Ensure that jobs already accepted by the ministry are not considered.
3. **NLP Process:** Use a machine learning model (SBERT) for natural language processing (NLP) to classify jobs into three categories: "already accepted," "rejected," and "new proposed jobs."
4. **Hybrid Approach:** After AI filtering, a committee (لجنة) would review the results and make the final decisions.
5. **User Interface:** Provide an intuitive website for administrators to upload CSV files, download filtered results, and facilitate committee review.

---

2. Project Workflow

Step 1: Data Collection and Preparation

- **Input Data:** The project involved working with multiple CSV files:
    1. **Activité-2025-02-18(1):** Contains jobs already accepted by the ministry.
    2. **Bita9a_Mihanya(2):** Contains jobs requiring technical approval ("Bita9a Fanya").
    3. **Nomenclature_column3_cleaned(2):** Contains jobs requiring a "Registre de Commerce."
    4. **Testing Data:** A file with 3,000 instances labeled as "validée," "N'est pas conforme," "à vérifier," "Déjà existante," or "manque de précision."
- **Data Cleaning:**
    - Removed duplicates and irrelevant entries.
    - Separated the 125 newly accepted jobs from the main dataset for testing purposes.

Step 2: Web Scraping

- **Live Web Scraping:** An automated system was proposed to keep these datasets updated in real-time.

Step 3: NLP Process (SBERT)

- **Model Selection:** SBERT (Sentence-BERT) was chosen for its ability to handle semantic similarity and NLP tasks.
- **Training Data:** The model was trained using:
    - The 125 newly accepted jobs.
    - Instances labeled as "Déjà existante" (already existing jobs).
- **Testing Data:** The model was tested on a dataset of 3,000 instances labeled by a single human annotator.

Step 4: Model Performance Evaluation

- **Initial Accuracy:** The model achieved an accuracy of 0.36 with a threshold of 0.5.
- **Challenges Identified:**
    - The testing data was labeled by a single person, leading to potential errors and misclassifications.
    - The low accuracy suggested the need for better training data or model adjustments.
- **Solution:** Focused on improving the training dataset by using the 125 newly accepted jobs and "Déjà existante" instances.

Step 5: Filtering and Classification

- The system processed 15,000 instances and generated three output files:
    1. **Already Accepted Jobs:** Jobs that are already accepted by the ministry.
    2. **Rejected Jobs:** Jobs requiring "Registre de Commerce" or "Bita9a Fanya."
    3. **New Proposed Jobs:** Jobs that meet the criteria for auto-entrepreneurial consideration.

Step 6: Committee Review (لجنة)

- After the AI filtering process, the results are reviewed by a committee (لجنة) composed of administrators and experts.
- **Responsibilities of the Committee:**
    - Review the AI-filtered results.
    - Make final decisions on which jobs to accept or reject.
    - Ensure that the decisions align with the ministry's policies and objectives.
- **Hybrid Approach:** Combines the efficiency of AI with the expertise and judgment of human reviewers.

Step 7: Website Development

- **Frontend:** Built using React for a responsive and user-friendly interface.
- **Backend:** Developed using Flask to handle file uploads, data processing, and model integration.

- **Functionality:**
    - Administrators upload a CSV file.
    - The system processes the file and generates three output CSV files for download.
    - The committee (لجنة) can access the filtered results through the website for review and decision-making.

---

3. Challenges and Solutions

Challenge 1: Low Model Accuracy

- **Issue:** The SBERT model achieved an accuracy of only 0.36 on the testing set.
- **Root Cause:** The testing data was labeled by a single person, leading to potential inconsistencies.
- **Solution:** Improved the training dataset by using the 125 newly accepted jobs and "Déjà existante" instances. This provided a more reliable foundation for the model.

Challenge 2: Data Updates

- **Issue:** The datasets for jobs requiring "Registre de Commerce" and "Bita9a Fanya" needed to be updated regularly.
- **Solution:** Implemented web scraping to dynamically fetch and update these datasets.

Challenge 3: Hybrid Approach Implementation

- **Issue:** Ensuring a smooth workflow between AI filtering and committee review.
- **Solution:** Designed the website to facilitate seamless interaction between the AI system and the committee, allowing for easy review and decision-making.

---

4. Results and Outcomes

- **Model Performance:** After refining the training data, the SBERT model showed improved performance in classifying jobs.
- **Output Files:** The system successfully generated three CSV files:
    1. Already accepted jobs.
    2. Rejected jobs.
    3. New proposed jobs.
- **Committee Review:** The hybrid approach ensured that final decisions were made with both AI efficiency and human expertise.
- **Website Functionality:** The website allowed administrators to upload CSV files, download filtered results, and facilitated committee review.

---

5. Future Improvements

1. **Model Enhancement:** Explore other machine learning models or fine-tune SBERT further to improve accuracy.
2. **Automated Updates:** Implement live web scraping to keep datasets updated in real-time.
3. **User Feedback:** Incorporate a feedback mechanism for administrators and the committee to report errors or suggest improvements.
4. **Scalability:** Optimize the backend to handle larger datasets and more complex queries.
5. **Committee Dashboard:** Develop a dedicated dashboard for the committee to streamline the review process.

---

6. Conclusion

The project successfully delivered a website for Algeria Venture's administration, enabling efficient filtering and classification of proposed jobs and projects. By leveraging SBERT for NLP, web scraping, and a hybrid approach involving a committee (لجنة), the system addressed the ministry's requirements while providing a user-friendly interface for administrators. Despite initial challenges with model accuracy, the team implemented effective solutions to ensure reliable results. This project lays the foundation for future enhancements and scalability, ensuring a balance between automation and human judgment.

ù