|  |  |
| --- | --- |
| Resume Parser Project: Meeting minutes | |
| PROJECT: | Building a Resume Parser in Python using NLP and deploying the model on a cloud based platform using an application. |
| DATE: | February 1 – February 24, 2023 |
| MEMBERS: | SIDDHARTHA PATRA (0787887) |
|  | USAMA NORAT (0792225) |
|  | AMIR DAHYA (07912521) |

# Specific Activities from the prior 4 weeks

* **Model evaluation**: After creating baseline model, we evaluated the trained model's performance based on some new unknown resumes.
* **Debugging and testing**: We had to go through several iterations, including debugging, testing, and fixing issues encountered during development. This step was crucial to evaluate the model's performance in real-world scenarios and identify any areas for improvement.
* **Code review**: We conducted a thorough code review to identify potential issues and improve the code quality. This step helped us identify areas where the code could be optimized, increased performance, and enhanced the overall quality of the code.
* **Research and implementation**: We researched and started implemented the usage of Google Cloud Platform's engine APIs to deploy the code for the baseline model on the platform using the Streamlit app. This involved exploring various options, reading documentation, and finally integrating the APIs with our code.

# Specific Outcome from prior 4 weeks

* 1. In the project code development phase, we debugged, tested, and made some changes, resulting in a more robust and efficient NLP pipeline.
  2. Researched Google Cloud Platform's engine APIs for ML model deployment and identified the components required for ML model deployment on GCP.
  3. Learned about Cloud Source Repository and how to create a new repository or connect to an external repository, also clone the repository, and upload necessary files to the repository in zip format. and how to unzip the folder and move to the git folder and follow git commands to add the files to the Cloud Source Repository.
  4. Learned about Cloud Build Trigger and how to create a new cloud build trigger, how to run the trigger and create a docker file.
  5. Learned about Google Container Repository (GCR) and how to deploy images to it.
  6. Learned how to create a VM instance for the virtual machine, how to connect to the VM instance.
  7. Learned how to test the baseline model on real-world data to check for NER accuracy.
  8. Continue exploring the Google Cloud Platform to gain more knowledge and expertise in ML model deployment and Learn about the different tools and resources available on GCP for ML model deployment.
  9. Begin implementing GCP's engine APIs to deploy the code for the baseline model on the platform using the Streamlit app.

# On Target

The project is on target and the status is green, as we have successfully developed our baseline model using the SpaCy library to perform entity recognition and identify key entities in a resume. We were also able to successfully test our model on new unseen data (resumes) and check the accuracy of our entity identification model. And we researched and learned about GCP deployment and started to work on deployment.

# Challenges/Disagreements

1. Technical difficulties: The process of learning and implementing new technologies can be challenging and time-consuming, and for some concepts we encounter technical difficulties that require additional troubleshooting.
2. Complexity of the platform: Google Cloud Platform is a complex platform with many different tools and services, and it is overwhelming for us to navigate and understand all the different components required for ML model deployment.
3. Disagreements on implementation: As the team begins to implement the deployment process, there may be disagreements on the best way to approach certain tasks or on the best tools to use.
4. Time constraints: The team have deadlines that make it challenging to fully explore and understand all of the components of GCP required for ML model deployment.

# Planned Activities for coming weeks

The team has planned the following activities for the coming week:

* + - Research and explore more about Streamlit app deployment on Google Cloud Platform.
    - Set up a virtual environment and install necessary dependencies for Streamlit app development.
    - Write Streamlit app code and integrate it with the baseline model code.
    - Configure Cloud Build Trigger to automate the deployment process and build/deploy the Streamlit app on GCP.
    - Conduct tests to ensure that the deployment is successful and make any necessary adjustments.

These planned activities will keep the project on track and bring us closer to delivering a functional Resume Parser in Python using NLP and deploying it on a cloud-based platform.