Project Title: Named Entity Recognition

**Project Details:**

**DATASET:**

* Resume Entities for NER

( Link: <https://www.kaggle.com/datasets/dataturks/resume-entities-for-ner/data>).

* This dataset consists of 220 annotated resumes. I trained the model with 200 resume data and test it on 20 resume data.

**PREPROCESSING:**

* For this resume data, pre-processing involves removing extra whitespaces, tabs, and newlines and removing overlapping entities.
* And then converting it into a suitable format for training the spacy model.

**MODEL TRAINING:**

* Used python’s spaCy module for training the NER model.
* Loaded a blank English model and added the NER component to the model’s pipeline.
* The model is trained using a training loop that updates weights based on examples from the training data.
* This approach ensures that the model is trained effectively to recognize entities within resume data.

**PREDICTIONS:**

* The model is tested on 20 resumes and the predicted summarized resumes are displayed.

**Challenges & Solutions:**

* **Preprocessing:** Faced problem with handling overlapping entities as spacy will not accept overlapping while training the model.

**Solution:** I implemented a function to handle overlapping entities. This function removes or consolidates overlapping spans in the annotations, ensuring that each span is unique and does not overlap with others.

* **Model Training:** When training the spaCy model, faced a challenge to ensure that the data was correctly formatted and that the model was adequately tuned to handle the diverse nature of the resume data.

**Solution:** Ensured that the data fed into the model was in the correct format, including properly annotated examples and avoiding any data inconsistencies. Implemented regular evaluation checkpoints to monitor the model's performance and make necessary adjustments.

**References:**

* Kaggle: Collection of Dataset.
* DataCamp courses:

1. Introduction to natural language preprocessing in python
2. Natural language preprocessing with spaCy

* GitHub- <https://github.com/rijaraju/Resume-Parser>
* GitHub- [Entity-Recognition-In-Resumes-SpaCy/README.md at master · DataTurks-Engg/Entity-Recognition-In-Resumes-SpaCy (github.com)](https://github.com/DataTurks-Engg/Entity-Recognition-In-Resumes-SpaCy/blob/master/README.md)