Diary

Week 1

Week: 30-05-22 to 03-06-22

Major Learnings & Project Milestones:

1. About the PS1 Station: I was familiarized with couture.ai's work and ongoing projects as well as the work culture through an orientation as well as through stand-ups and training sessions.
2. Went through the training material provided by the company to get familiarised with the working and deployment of ML Models and understand the progress already made in the Project.
3. Learnt the basics of Python by YouTube tutorials and online resources. Practiced arrays and lists thoroughly due to their future relevance.
4. Finished the basic Pandas course on Kaggle along with practicing examples and implemented learnings on the History Dataset provided by the Industry Mentor.

Skills Gained (Both technical and soft skills):

1. Python
2. Pandas
3. Gathering resources relevant to the project effectively.
4. Observing the etiquettes and coordination among team members of the project team with seamless communication and frankness about problems faced during the stand up calls that are held every day.
5. Coordinating with a mentor and reporting progress everyday in the morning.

Major Challenges:

1. Was out of touch from coding for more than a year and learning a new coding language in a very short span of time.
2. Understanding the terms used by others in stand-up calls and training sessions due to no prior knowledge about ML concepts.

Week 2

Week: 06-06-22 to 10-06-22

Major Learnings & Project Milestones:

1. Learnt how to load, manipulate, and select data using numpy, as well as understand the fundamental data types in the numpy ecosystem. Also saw the benefits of vectorisation using numpy.
2. Finished the course 'Introduction to Data Science in Python' by University of Michigan on Coursera by auditing it.
3. Learnt how to manipulate strings more effectively using Regular Expressions and saw how regular expression pattern matching is expressed at a high level.
4. Set up Jupyter Notebooks and installed necessary libraries required for future use.
5. Went deeper into Pandas and learnt how to code more effectively and joining and analysing multiple datasets.
6. Analysed a considerably large dataset by joining History and Catalogue Data along a common index ('Product ID') and deriving inferences by considering both the information sets simultaneously.

Skills Gained (Both technical and soft skills):

1. Pandas
2. Regular Expressions
3. NumPy
4. Gathering resources relevant to the project effectively.
5. Clarifying doubts and asking for resources and extensions more openly when required from mentor.
6. Collaborating and sharing resources with my co-mentee.

Major Challenges:

1. Kept running into errors that would take a long time to debug.
2. Due to an information overload in a short span of time, started mixing up concepts and sometimes found it difficult to recall concepts.
3. While working with large datasets, Jupyter kept crashing and had to set up the environment on Google Colab instead. Run time was still extremely slow and had to try different syntaxes and methods to improve performance.
4. While using online resources, realised GeeksforGeeks had repetitive examples and had to browse for external better resources which eats up valuable time.

Week 3

Week: 13-06-22 to 17-06-22

Major Learnings & Project Milestones:

1. Finished the course 'Applied Machine Learning' by University of Michigan on Coursera by auditing it.
2. Learnt how to build and evaluated a basic k-nearest neighbours classifier on an example dataset using Python and scikit-learn
3. Learnt how to apply the right algorithm for a given task by understanding the strengths and weaknesses of specific supervised machine learning algorithms in Python with scikit-learn.
4. Learnt how to optimize a machine learning algorithm using a specific evaluation metric appropriate for a given task acknowledging how accuracy alone can be an inadequate metric for getting a complete picture of a classifier's performance.
5. Applied techniques like regularization, feature scaling, and cross-validation to avoid common pitfalls like under- and overfitting.
6. Learnt about important concepts like Data Leakage, Cross Validation, over and underfitting, etc.

Skills Gained or Concepts Learnt:

1. SciKit Learn
2. K Nearest Neighbours
3. Linear, Logistic, Lasso and Ridge Regression
4. Binary and Multiclass Classification
5. Kernelized Support Vector Machines
6. Cross Validation
7. Decision Trees
8. Evaluation Methods like precision, recall, etc.
9. Neural Networks

Major Challenges:

1. Lack of Practice Questions on concepts.
2. While the course did a good job at explaining all the concepts, it gave very little explanation on the coding aspect of it which took time to familiarise myself with.
3. Without having done Matplotlib, the code given to plot the regressions, etc for better visualisation was hard to understand.

Week 4

Week: 20-06-22 to 24-06-22

Learning Outcomes of the Week:

1. Understood basic hyperparameter concepts
2. Finished learning basic Machine Learning concepts

Skills Gained or Concepts Learnt:

1. Hyperparameters
2. NER

Major Learnings & Project Milestones:

1. Finished learning basic Machine Learning concepts
2. Three basic alogrithms used for hyperparameter tuning: Grid Search, Bayesian Optimisation and Random Search

Major Challenges:

1. Fell sick and had to take a sick leave for 3 days. Fell behind on progress.

Week 5

Week: 27-06-22 to 01-07-22

Major Learnings & Project Milestones:

1. Implemented and coded two major Hyperparameter Algorithms: Grid Search and Random Search.
2. Got started with Natural Language Processing concepts and got a general idea about NER using Spacey.
3. Got accustomed to coding using the Spacey Library.
4. Understood the implementation of Word2Vec using Python and the theory behind it along with the various different ways its done (CBOW and Skim Gram).
5. Got comfortable with implementation of Hyperparameter tuning algorithms by applying it to a Random Forest Model and a Gradient Boosted Classifier model.

Skills Gained or Concepts Learnt:

1. Hyperparameters
2. Grid Search
3. Random Search
4. Spacey
5. NER
6. Word2Vec (CBOW and Skip Gram)
7. Pipeline
8. Entity Ruler
9. Matcher

Major Challenges:

1. Implementing Hyperparameter Tuning proved to be tricky and I made a lot of basic errors that showed a lack of practice in basic concepts.
2. Finding Word2Vec Models that don't use TensorFlow, Keras or other libraries I'm not yet accustomed to.
3. Understanding Word2Vec Model was difficult as it used some higher level concepts.

Week 6

Week: 04-07-22 to 08-07-22

Major Learnings & Project Milestones:

1. Tried to implement two Hyperparameter Algorithms: Random Search and HyperBand using KerasTuner on the Categorization Output DL Model already developed by other team members.

2. Got familiarized with Deep Learning Concepts and dug deeper in Neural Networks like Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN).
3. Researched on various DL Model Layers like Dense, LSTM, etc.
4. Read up on Data Leakage and Overfitting, underfitting to avoid making similar errors while working with a real model.

Skills Gained or Concepts Learnt:

1. HyperBand Optimization
2. KerasTuner
3. CNN and RNN
4. Dense and LSTM Layer
5. Data Leakage

Major Challenges:

1. Implementing Hyperparameter Tuning proved to be tricky and I made a lot of basic errors due to being unexperienced with Data Learning Models.
2. Run times on the model were very high due to large datasets being processed.

Week 7

Week: 11-07-22 to 15-07-22

Major Learnings & Project Milestones:

1. Continued working on Hyperparameter Tuning of Categorization Model.
2. Started working on Autocompletion Model which made use of pretrained models from SBERT, which I tested to use the best one possible for the model.
3. Cleaned up the data and changed the approach used for word embeddings and Semantic Search to optimize the Autocompletion Model.

Skills Gained or Concepts Learnt:

1. BERT
2. Semantic Search
3. Word Embeddings

Major Challenges:

1. Hyperparameter Tuning the model proved to be difficult and I kept running into errors without knowing how to go about fixing them.
2. Autocompletion Model made use of very large datasets because of which computational run times were very large.

Week 8

Week: 18-07-22 to 22-07-22

Major Learnings & Project Milestones:

1. Continued working on Autocompletion Model.
2. To improve recall tried doing prefix extraction and experimented by tuning various parameters in the model.
3. Tried using Levenshtein Distance and Similarity and combined the bert score with the Levenshtein Score to reorder generated queries and produce better results to increase recall.

Skills Gained or Concepts Learnt:

1. BERT
2. Prefix Extraction
3. Levenshtein Distance