

Berichterstattung über das Praktikum

Die Berichterstattung über Ihr Praktikum hilft anderen Studierenden dabei, sich einen Eindruck davon zu machen, welche Tätigkeiten Sie in den jeweiligen Betrieben ausüben durften, und wie Sie die Praktikumszeit erlebt haben und beurteilen.

Unternehmen

| | | | |
|--------------|-------------------------------|-----------|--------|
| Unternehmen: | Spiced Academy | Branche: | Berlin |
| Dauer: | 12 weeks | Standort: | Berlin |
| Studiengang: | Masters, Scientific Computing | | |

Bewerbung

Art der Bewerbung:

Email

Warum haben Sie sich für dieses Unternehmen entschieden?

I took interest in Data Analytics/Science after taking a course 'Introduction to Data Analytics with R' and so I decided to apply for an intern role as a Data Analyst. However, I was asked for more work experience. So I decided to enrol for the 12 weeks intensive bootcamp at Spiced Academy to arm me with my skills as a Data Scientist.

Tätigkeiten und Bewertung

Beschreibung der Tätigkeiten:

Every week, we worked on projects and present every friday based on what was taught and learnt during the week. I learnt different concept of MACHINE learning and being a Data Scientist.

Week 1: This week's project was about creating a git repository where we pull our lecture notes from and also where we submit (push) our daily task to. We learnt how to import a dataset from a csv (and other format) into a python environment, especially ipython also known as Jupyter notebook. We also did data wrangling, how to analyze a dataset and also how to visualize it and make it understandable for stakeholders to understand. The project was based on analysis, creating different graphs such as scatterplot, histogram using Matplotlib.

Week 2 : I used machine learning models to predict passengers' survival on the Titanic. This is a classification problem and so we used the machine learning model for classification which is Logistic Regression, decision trees and random Forest. We calculated the train and test accuracy and also did cross validation score to test the models validity. We also calculated the loss function in Logistic Regression and also evaluated our classifiers.

Week 3: The goal for this week's project is to build and train a regression model on the Capital Bike Share (Washington, D.C.) Kaggle data set, in order to predict demand for bicycle rentals at any given hour, based on time and weather, e.g. "Given the forecasted weather conditions, how many bicycles can we expect to be rented out (city-wide) this Saturday at 2pm?"

I splitted the data into a training and test set, conducted an exploratory data analysis, trained a regression model, iteratively optimized the model by expanding or selecting features, calculated a RMSLE for the training and test set and then uploaded my code to GitHub.

Week 4: In this week's project, I built a text classification model on song lyrics. The task is to predict the artist from a piece of text. To train such a model, I first collected my own lyrics dataset. I used selected albums from BlackEye Peas and 50 cents. I downloaded the html page which has the songs by web scraping it and using regular expressions, I then extracted the hyperlinks of the songs page. Since we need the lyrics of the songs, I vectorized the words using the bag of words method, then trained a classification model that predicts the artist from a piece of text.

Week 5: In this week's project, I build a dashboard summarizing the Northwind Database. It is a sample database that is shipped along with Microsoft Access. The data is about "Northwind Traders", a fictional company. The database contains all sales transactions between the company and its customers as well as purchases from Northwinds suppliers. I also learnt about using databases such as PostgreSQL and how to access the data stored in it.

Week 6: In this project, I built a data pipeline that collects tweets and stores them in a database. Next, the sentiment of tweets is analyzed and the annotated tweets are stored in a second database. Finally, the best or worst sentiment for a given is published on Slack every 10 minutes.

Week 7: I got the temperature data from www.ecad.eu. I then built a baseline modeling the trend and seasonality of the dataset. I plotted and inspected the different components of the time series and then removed the trends and seasonality from the dataset. Then I modeled the remainder after the trends and seasonality was removed using the AutoRegression model. Then I tested the remainder for stationality.

Week 8: This week, I wrote a program that simulates customer behavior in a supermarket using Markov Simulation. This will help to understand our customers better in order to optimize the layout, staffing and service of our supermarkets. I modeled the way customers move through a representative shop. The main business goals are to understand customer behavior, to explain customer behavior to our non-data staff and to optimize staffing so that the queues do not get unnecessarily long. I used the following to model a supermarket with six areas: entrance, fruit, spices, dairy, drinks and checkout.

Week 9: The goal of this week is to build an Artificial Neural Network that recognizes objects that you hold into the webcam. I did so by collecting raw data of oranges and apples from my webcam, built a neural network from scratch, and then trained the neural network using the dataset that I have. I also learnt about pre-trained neural networks and transfer learning.

Week 10: This week's goal was to build a recommender system with a web interface and I used a movies dataset. My recommender system recommends movies based on what you like or have watched before.

Week 11: This week, I learnt how to apply software engineering to my movie recommender. This includes learning how to build a webpage with HTML, CSS and Javascript. I also learnt how to use streamlit to build my app which is what I used for my final project at the end of the bootcamp.

Week 12: This week was all about working on final project which we presented to everyone who attend. We had to do an online graduation due to the COVID restrictions. I built a recommender app to recommend Youtube videos similar to the ones a viewer has seen before. I used the Trending videos API to get the trending videos from 45 countries all over the world. I analyzed the data, drew some graphs and then went ahead to build an app that recommends similar videos based on what the viewer has watched and likes.

In all, I am glad I enrolled in the bootcamp as it has opened my eyes to different possibilities. I am well equipped and would also like to implement this in my thesis as would be researching (starting in January) Random Forests which next to deep learning are one of the most powerful machine learning approaches.

Wie würden Sie Ihren Praktikumsplatz bewerten und was ist Ihr Fazit?

I loved it and would recommend to anyone.