|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Time Sheet for: | | | | | | <Your name here> | |
| Covering time period: | | Sept   Jan  | | | Oct   Feb  | | Nov   Mar  |
| Date: | Start Time: | | End Time: | Description: | | | |
| 05.11 | 8:00 | | 8:20 | * Presentation of the previous project | | | |
| 05.11 | 8:20 | | 9:00 | * Setting up working environment   + Python   + Anaconda   + Pydotplus   + Tensorflow   + Pandas library   + NumPy   + Udemy Jupiter course materials | | | |
| 05.11 | 9:00 | | 9:20 | * Exploring Jupiter examples * Testing Pandas library, displaying CSV files as tables and manipulating them | | | |
| 05.11 | 9:20 | | 9:35 | * Types of data research (numerical, categorical, ordinal) * Recap of mean, median, mode * Trying it in python | | | |
| 05.11 | 9:35 | | 9:50 | * Variation, population and sample variance, standard deviation research (used to identify outliers) * Trying it in python | | | |
| 05.11 | 9:50 | | 10:00 | * Probability Density (continuous data) & Probability Mass (discrete data) research | | | |
| 05.11 | 10:00 | | 10:15 | * Steve talk about creating custom circuit boards | | | |
| 05.11 | 10:15 | | 10:25 | * More data distributions research (Uniform, Normal / Gaussian, Exponential, Binomial & Poisson Probability Mass) | | | |
| 05.11 | 10:25 | | 10:40 | * Percentiles and Moments (mean, variance, skew, kurtosis) * Trying it in python | | | |
| 05.11 | 10:40 | | 11:15 | * More research on MatPlotLib to create better looking graphs (multiple plots, adjust axes, ticks, grid, add legend, plot pie & bar chart, scatter plot, box & whisker, save to file) | | | |
| 07.11 | 10:00 | | 10:50 | * Another graphs library research – Seaborn * Advanced visualization | | | |
| 07.11 | 10:50 | | 11:25 | * Covariance & Correlation understanding * Research on ways to measure it * Coding it hard way in Python and using NumPy functions | | | |
| 07.11 | 11:25 | | 12:00 | * Conditional Probability recap * Ways to compute * Creating fake data in Python and practicing this concept | | | |
| 14.11 | 9:00 | | 9:30 | * Studying Bayes Theorem * Understanding its significance * “The probability of something given something else is not the same thing as the other way around” | | | |
| 14.11 | 9:30 | | 9:50 | * Starting with Regression Analysis * Linear Regression research * Practicing this concept in Python | | | |
| 14.11 | 10:00 | | 10:25 | * Polynomial Regression research * Trying the concept in Python (PyPlot function is amazing) | | | |
| 14.11 | 10:30 | | 11:00 | * Multiple Regression research * Standard notations in Machine Learning (X – data we encounter; y – something we try to predict) * Statsmodels library practice, helps to compute OLS Regression and print meaningful results for analysis * Using this data to make an actual prediction by scaling my desired variables for prediction into the same scale used to train the model | | | |
| 19.11 | 8:00 | | 8:20 | * Research on multi-level models, understanding the complexity and use of the concept, setting some goals for future (the topic is way beyond the scope of this project) | | | |
| 19.11 | 8:20 | | 9:00 | * General research on Supervised and Unsupervised Machine Learning * Research on Train/Test supervised learning concept * Introduction to K-fold Cross Validation (supervised machine learning approach) * Using Train/Test in Python to prevent overfitting a polynomial regression | | | |
| 19.11 | 9:00 | | 9:40 | * Helping out the classmate with his Python problem * Python IDLE could not find the module, even though it was installed, turned out the IDLE was using different Python version from the one where the required module was installed, so had to figure out how to install the module to the right place. Everything works now. | | | |
| 19.11 | 9:50 | | 11:00 | * Bayesian Methods research (making use of Bayes Theorem for machine learning) * Understanding how it is applied in spam detection systems * Assumption that the presence of different words are independent of each other is reason why the concept is called “Naïve Bayes” * Scikit-learn library makes this process pretty easy * Implementing a Spam Classifier in Python using Scikit-learn | | | |
| 21.11 | 10:00 | | 10:30 | * Steve talk about not getting shocked/electrocuted | | | |
| 21.11 | 10:30 | | 10:50 | * Continue to experiment with Spam Classifier * Example looks and works great, but how smart it actually is? * Tried splitting training data into train/test subsets and trying the model on numerous unknown spam mails (now classifier fails quite often – 45% fail on 500 test data taken 2400 training data) | | | |
| 21.11 | 11:00 | | 11:50 | * K-Means Clustering research (common unsupervised learning technique) * Practicing this technique in Python using Scikit-learn | | | |
| 26.11 | 8:05 | | 8:15 | * Understanding entropy and the ways to measure it * Prerequisite for Decision Trees study | | | |
| 26.11 | 8:15 | | 9:15 | * Decision Trees Concepts research (supervised learning technique) * Practicing this technique in Python, creating decision trees flowcharts * Problem: GraphViz's executables not found * Solution: did not find a good reason for the problem, however fixed it by modifying source code of the library (hardcoded path to GraphViz installation folder) | | | |
| 26.11 | 9:20 | | 10:00 | * Random Forest research – example of Ensemble Learning (multiple decision trees for collective predictions) * Bootstrap aggregating (used by random forests) * Boosting – enhancing mis-classified attributes * Bucket of models – several different models * Stacking – run multiple models at once and combine the results | | | |
| 26.11 | 10:00 | | 10:50 | * Support Vector Machines research (advanced mathematical supervised techniques for classifying higher-dimensional data) * Practicing this in Python with Scikit-learn | | | |
| 28.11 | 10:10 | | 10:40 | * Starting with research on recommender systems * User-Based Collaborative Filtering technique research * Understanding its principles and problems with it (e.g shilling attack) | | | |
| 28.11 | 10:45 | | 11:15 | * Item-Based Collaborative Filtering technique research * It addresses a lot of shortcomings of user-based collaborative filtering | | | |
| 28.11 | 11:15 | | 11:50 | * Applying Item-Based Collaborative Filtering in Python * Starting to work on a movie recommender system * Working on script that identifies movie similarity based on real movie ratings data provided by GroupLens (first step of item-based collaborative filtering) | | | |
| 03.12 | 8:10 | | 8:50 | * Continue working on a movie recommender system * Analyzing first results and improving the system | | | |
| 03.12 | 8:50 | | 9:40 | * Understanding why the results were poor and how they could be improved * Experimenting how different variables can affect the result | | | |
| 03.12 | 9:40 | | 11:10 | * Working on movie recommender system * Experimenting with different variables * Adding my own movie rating data to give recommendations based on that * Filtering the results to be more readable and useful | | | |
| 03.12 | 18:20 | | 20:00 | * Experimenting with the code for recommender system trying to get better results * Research ways of improvement recommender system logic * Research causes of bad results | | | |
| 03.12 | 20:00 | | 21:20 | * Working on improvement for recommender system * Adding new concept to the logic of recommender system: * Identifying movies similar to the ones that user rated low and penalyzing them in order to produce better results | | | |