

Ammar Alyousfi



KL, Malaysia • ammar5656@gmail.com • ammar-alyousfi.com • +601162078955

Education

Master of Data Science at University of Malaya, Malaysia 2018 → (Sep 2019, expected)

- Relevant courses: Data Analytics, Programming for Data Science, Big Data Application and Analytics, Big Data Management, Data Mining, Principles of Data Science, Machine Learning for Data Science.

B.Sc in Computer Engineering at Princess Sumaya University for Technology, Jordan 2012 → 2016

- Got the first rank in the Computer Engineering department with a GPA of 93.9%
- The program is accredited by ABET.
- Received an 'Outstanding Student Award' for academic achievement in 2015 and other awards.
- Relevant courses: Data Structures and Introduction to Algorithms, Database Systems, Computer Architecture and Organization, Visual Programming, Discrete Mathematics, Calculus 3, Operating Systems, Signals and Systems, etc.

Skills

Programming Languages: Python, JavaScript, limited experience in R.

Data analysis and machine learning: Pandas, Numpy, Scikit-learn, Jupyter Notebook, SQL, Amazon Web Services (EC2, S3), limited experience in Hadoop.

Text analysis and NLP: Regular expressions, Wordcloud, limited experience in NLTK.

Web scraping: BeautifulSoup, Selenium.

Data visualization: Matplotlib, Seaborn, limited experience in Tableau.

Web development: Frontend (HTML, CSS, JavaScript, Vue.js), Backend (Django), hosting and deployment.

Misc: Lyx (LaTeX IDE), Git, graphic design (Adobe Photoshop, GIMP), Google Analytics.

Languages: fluent in English (TOEFL: 102/120 in 2016), native Arabic, basic Malay.

Projects

YouTube Trending Video Analysis: Analyzed a dataset describing 40,000+ trending YouTube videos to answer interesting questions like "What are the most common words in video titles?", "Which channels and which categories have the largest number of trending videos?", etc. Also looked at the relationship between likes and views, correlation between variables, the publishing time distribution of trending videos, and many more. (Python, Pandas, Matplotlib, Seaborn, WordCloud, Jupyter Notebook).

Analysis of Top Reddit posts: Analyzed the top 1,000 posts of 18 popular subreddits on reddit.com. Found the most common words and n-grams, used word clouds to show them, explored title-length and comment-count distribution, etc. (Python, Pandas, NLTK, WordCloud, Matplotlib, Seaborn).

Kaggle Machine-Learning Competitions: Participated in many competitions. Some are:

- Help Navigate Robots: (best score: 1.0000, worst score: 0.0000, my score: 0.6802 [top 5%], 1478 competitors): Participants were asked to detect the type of the surface the robots are standing on using data collected from Inertial Measurement Units (IMU) sensors. Dealt with ~1 million rows of time-series data; applied feature extraction using tsfresh library; used Microsoft's LightGBM for modeling. Evaluation metric: Multiclass Accuracy.
- PUBG Finish Placement Prediction (best score: 0.01385, worst score: 0.52662, my score: 0.04227, 1780 competitors): Participants were asked to predict players final placement using final in-game stats and initial player ratings. Dealt with a 6+ million rows of data; applied extensive feature engineering; used XGBoost for modeling. Evaluation metric: MAE.
- VSB Power Line Fault Detection: (best score: 0.71899, worst score: -0.28109, my score: 0.58655, 1595 competitors): Participants were asked to detect partial discharge patterns in time-series signals acquired from overhead power lines. Dealt with a 10+ GB of data; applied signal filtering, feature extraction and selection; used stacking for modeling with many models: LightGBM, neural net, KNN, etc. Evaluation metric: Matthews correlation coefficient.

House Price Prediction, An End-to-End ML Project: End-to-end analysis including data preparation and cleaning, exploratory data analysis, feature engineering and hyperparameter tuning, modeling using KNN, SVM, neural networks, and other models, and evaluation and result analysis. (Python, Pandas, Matplotlib, Seaborn, Scikit-learn, XGBoost).

s3upload: Automates uploading files to AWS S3. (Python, boto3).

Focus Phase: An open-source time-tracker command-line application with statistics and visualizations. Published on the Python Package Index. (Python, process management, Matplotlib).

My personal website: Contains my portfolio, resume, and blog. (frontend and backend development, deployment, Python, Django, JavaScript, AWS, HTML, CSS).

Pair & Compare: Makes it easier for developers to choose the best fonts and font-pairs for their projects. It allows using all Google font without downloading or installing any of them. (Vue.js, JavaScript, HTML, CSS, Web Font Loader).

To view these projects and more, please visit my personal website (amm-alyousfi.com), Kaggle profile (kaggle.com/amm111/kernels), Github profile (github.com/amm1y), and Stack Overflow profile (stackoverflow.com/users/2282785/amm1y).