Ammar Alyousfi

Location: Dubai, United Arab Emirates
Email: ammar5656@gmail.com
Mobile: +971509524537 • WhatsApp: +601162078955

Github: ammar1y
Kaggle: ammar111
LinkedIn: ammar-alyousfi
Stack Overflow: 2282785
Website/blog: ammar-alyousfi.com

Education:

University of Malaya — Kuala Lumpur, Malaysia
Master of Data Science degree with distinction (GPA: 3.98/4) — Feb 2018 - Oct 2019

- Research project: built and deployed a machine-learning system to predict taxi-trip duration. Tools used include Python, Pandas, Matplotlib, Seaborn, Scikit-learn, XGBoost, Flask (for deployment), and Jupyter Notebook. Project code and report are available on `http://bit.ly/ds-prj`.
- Relevant courses: Data Analytics, Data Mining, Machine Learning for Data Science, Programming for Data Science, Big Data Application and Analytics, Big Data Management.

Princess Sumaya University for Technology — Amman, Jordan
B. Sc. degree in Computer Engineering; first rank (GPA: 93.9%) — Sep 2012 - Sep 2016

- Relevant courses: Data Structures and Introduction to Algorithms, Database Systems, Visual Programming, Discrete Mathematics, Computer Architecture and Organization, Operating Systems.

Experience:

Apigate Sdn Bhd — Kuala Lumpur, Malaysia
Data Science and Analytics Intern — April 2019 - July 2019

- Performed extensive data analysis on millions of data records stored in Google BigQuery to get useful business insights. Tools used include SQL, Python, Pandas, Google Data Studio, and Matplotlib.
- Created an ETL Python script to transfer and sync company data from Salesforce to Google BigQuery. The program was scheduled to run weekly and used API services to access Salesforce and BigQuery data.

Skills:

Programming Languages: Python, R, JavaScript.

Data Analysis and Machine Learning: Pandas, NumPy, Scikit-learn, TensorFlow/Keras, XGBoost, LightGBM, etc. Also familiar with SAS and Hadoop ecosystem.

Data Visualization and Dashboards: Matplotlib, Seaborn, Google Data Studio, Follium for maps, etc.

Databases: SQL. **Web Scraping:** BeautifulSoup, Selenium, HTTrack.

NLP: NLTK. **Web Development:** Flask, Django, HTML, CSS, JavaScript, Vue.js.

Cloud Computing: Google Cloud Platform (BigQuery for big data, Compute Engine, and Storage), Amazon Web Services (EC2, S3, and Route 53).

Languages: English: fluent (TOEFL iBT: 102). Arabic: native.

Projects:

YouTube Trending Videos Analysis: Analyzed data of 40,000+ YouTube trending videos to identify common patterns and get insights (Python, Pandas, Seaborn, etc.). Code and results: `http://bit.ly/YT-analysis`.

End-to-end data-science projects: Built a machine-learning system to predict house prices based on many characteristics like house size, construction year, etc. Project report and code: `http://bit.ly/hp-pdf`.

Kaggle Competitions: Participated in many machine-learning competitions on Kaggle. Some are:

- **Help Navigate Robots:** The goal was to detect the type of surface a robot is standing on using sensor data. Ranked in the top 5% among 1470+ competitors. Used Python, LightGBM for modeling, and Tsfresh package to extract features from the time-series data. Code and results: `http://bit.ly/help-robo`.
- **VSB Power Line Fault Detection:** The goal was to detect whether power-line signals are faulty or not. Used Python, Pandas, NumPy, and SciPy. Models used include XGBoost, LightGBM, and Neural Network. Code and results: `http://bit.ly/vsb-pl`.

Clustering and Comparing the Neighborhoods of New York City and Toronto: Clustering was based on the similarity between the venues in the neighborhoods. Foursquare API was utilized to retrieve venues data. Project page: `http://bit.ly/clustnt`.

Pair & Compare: A web application that makes it easier to compare fonts and font-pairs. All 800+ Google fonts can be used without downloading or installing any of them. It is built using HTML, CSS, JavaScript, Vue.js, etc. It can be visited on `http://bit.ly/p-and-c`.

Focus Phase: An open-source time-tracking command-line application with statistics and visualizations. It is built using Python and published on the Python Package Index. Github link: `http://bit.ly/focus-phase`.

S3upload: An open-source Python application that makes it faster to upload a large number of files to AWS S3. Github link: `http://bit.ly/s3upload`.

Analysis of Stock Prices in Malaysia: Stock-prices data for 1800+ companies was crawled from many sources for 3 months; then multiple analyses were applied including sentiment analysis, stock-prices correlation, investment recommendation, and clustering. Github (with PDF report): `http://bit.ly/dm-assign`; video: `http://bit.ly/dm-vid`.

Certifications:

IBM Data Science Professional Certificate (2019)

- Courses include: Data Analysis with Python, Databases and SQL for Data Science, Data Science Methodology, Machine Learning with Python, Data Visualization with Python, and Open Source tools for Data Science. More info here: `http://bit.ly/IBM-ds`