Marc Nguyen

Tel: 0450519972
E-mail: marc.n.224@gmail.com

Linkedin:
https://www.linkedin.com/in/nguyen-marc/

Personal statement

I am currently completing my Master of Data Science at The University of Melbourne, and hold a Master of Teaching (Secondary) from the Melbourne Graduate School of Education. I find data science to be an interesting combination of mathematical understanding, technical/analytical skill, and communication. I believe that my prior learning and employment has provided me with a good foundation for mathematical understanding, my current university course and personal projects have enriched my technical skills, and my teaching experience and university projects have supplied ample experience in effective communication and collaboration. I am interested in utilising data to explain behavioural patterns and inform decisions, yet understand that data is not the be all and end all. As Hans Rolling said: "The world cannot be understood without numbers. But the world cannot be understood with numbers alone".

Technical skills

- Python
  - Confident with
    - NumPy, pandas, sklearn, Plotly, Dash, TextBlob, BeautifulSoup, PyTorch, TensorFlow
  - Have used and am still developing my skills with
    - Django, GeoPandas, Shapely, Paramiko, HoloViews, Bokeh, Google API, mpi4py, NLTK, OpenCV, PyTesseract
- Databases
  - Building physical database models based on given business needs, and constructing appropriate SQL queries
  - Utilising CouchDB to store data, and creating views using MapReduce to streamline analysis
- R
  - Experience building generalized linear models and using goodness-of-fit tests to determine their appropriateness
- Weka
  - Using Machine Learning to build models which classifies the location of tweets
- Tableau (Public visualisations available at https://public.tableau.com/profile/marc.nguyen#!/
  - Confident in multiple forms of data visualisation and linking multiple data sources

Selection of projects

- Victorian Real Estate Data (Personal)
  - Scraping data from public sources on the internet using BeautifulSoup and requests
  - Building a Tableau dashboard to visualise trends in Victorian real estate data
    - To better inform potential buyers and sellers of up to date price trends by suburb
  - Geospatial feature engineering to provide more data using the google maps api and geopandas
  - Modelling to predict house prices using deep neural networks and TensorFlow
  - Key insights
    - While market sentiment is down, there are many suburbs with continued growth
    - Overall price direction varies based on number of bedrooms
    - The pandemic led to a large shift from auctions to private sales
- Carlton Traffic Speed prediction (University - Data Science Industry Project with the Department of Transport)
  - Data extraction, feature engineering and exploratory data analysis using pandas
  - Trained and systematically compared a variety of models, including Linear, Neural Network, and LSTM in order to predict traffic speeds throughout time periods
  - Built a dashboard using plotly dash to present model predictions for user-input parameters
  - Key insights
    - Linear models are able to capture peaks and throughs of traffic speed, however time series models are better at capturing exceptionally high/low speed periods
    - Time of day, day of week and seasonality (regular, school holiday, public holiday) were key

- Victorian Covid-19 visualisation (Personal)
    - Building a Tableau dashboard to visualise new/current cases over time for individual LGAs, and analysing case numbers compared to demographic information
    - Key insights
        - When adjusting for population, there are no distinct trends between LGA education level, income, or population size and incidence of COVID-19

- Power Demand and Rate visualisation (2020 Melbourne Datathon)
    - Analysing data from AEMO to gain insight into the impact of the COVID-19 pandemic on power
    - Building a Tableau dashboard to visualise demand and rate by State, year, month, and day-of-week
    - Tuning and training linear models, DNNs and LSTMs to predict demand (Mean Absolute Percent Test Error ~ 6%)
    - Key insights
        - Demand during the off peak in Victoria is down from 2019, but the 5pm peaks in June/July were higher, potentially as a result of an increased proportion of people working from home.
        - COVID did not affect prediction ability by much, which has financial implications for individuals and businesses (as electricity costs shift from employers to employees)

Education and Qualifications

- University of Melbourne 2010-2013, Bachelor of Science (Applied Mathematics)
- Melbourne Graduate School of Education 2014-2015, Master of Teaching (Secondary Education)
- University of Melbourne 2019-2019, Graduate Diploma in Data Science (WAM: 82.6)
- University of Melbourne 2020-2021 Master of Data Science (scheduled completion July 2021) (Current WAM: 93)

Personal Information

Nationality: Australian citizen, (Vietnamese/Malaysian parentage)

Career Experience

Secondary School Teacher (January 2016-)

Siena College

- Teaching experience: Mathematics, Junior Science, Physics, Digital Technologies
- Collaborating within multiple faculties and creating/refining tests, exams, revision, topic outlines and professional learning
- Communicating with parents and providing feedback about student learning
- Adapting practice based on professional development and student needs

Intern (July 2020 – October 2020)

Melbourne Data Analytics Platform

- Integrated in a multidisciplinary team of 3rd space data professionals
- Built a web-dashboard (GLURMO) for live requesting SLURM data using sacct and presenting job statistics visualisations
    - Dynamically added and adjusted features in response to regular meetings and feedback
    - By utilising GLURMO, users of the SPARTAN cluster could obtain valuable information about their job resource usage and efficiencies

Research Assistant (July 2020 – December 2020)

Melbourne Data Analytics Platform

- Experience in project management for public intervention campaign
    - Worked across a variety of sub-teams to set up an efficient project structure in Microsoft Teams

Link to Melbourne Datathon 2020 Submission:

https://github.com/MarcNguyen22/MelbourneDatathon2020