Joe Tran

Mobile: +65 93495804 | **Email:** joetrankang@gmail.com
Blog: https://medium.com/@joetrankang | **GitHub:** https://github.com/joetrankang

I am a data scientist who is passionate about data and machine learning, especially NLP. I do have 2 certificates on NLP from Microsoft and DeepLearning.ai, and currently spend most of my free time finding some NLP projects to tackle and upgrade my skills. Please refer to page 3 for some of my projects.

I believe that with my skillset and more importantly, the right attitude, I can add value to your team. Should there be any specific skill/tech-stack required but not in my skillset, I will make sure that I learn and master it on the job.

SKILLS

- **Python:** Numpy, Pandas, Scikit-learn, Keras, Streamlit, Tensorflow, Flask
- **Database:** MySQL, PostgreSQL, MongoDB, Amazon Redshift and S3
- **NLP:** NLTK, TextBlob, Word2Vec, Bert, GloVe, Seq2seq
- **Web development:** HTML5/CSS3, JavaScript, React, Node.js, Express
- **Auto-ML:** H2O, PyCaret, AutoKeras, Auto-SKlearn, TPOT
- **Others**: R, Git, Tableau, Power BI, Amazon SageMaker
- **Strengths**: A/B testing, Experimental Design, ETL, Text Mining, Data Mining, Recommendation Engines, Customer Segmentation and Retention Models, Summarisation, Sentiment Analysis, Topic Modelling
- **Soft skills**: Stakeholders Management, Product Management

WORK EXPERIENCES

PERROMART | E-commerce

Data Scientist *May 2020 – present*

- Built recommendation engine in Python using past order history and deployed to production on Heroku, resulting in a 25% uplift in AOV
- Built a customer attrition random forest model that identify who are likely to churn and the root causes, hence lowering the churn rate by 7%
- Used tools such as AWS SageMaker (Python) to deploy trained models into production environment with minimal codes required by the developers
- Implemented a web crawler to collect training data for text classification models using Scrapy, Beautiful Soup and Python
- Collaborated with the marketing team to determine when the customer may come back using RFM model, therefore improving the overall retention rate by 5%
- Developed web-app with Streamlit and Python on AWS, displaying cohort retention and payback ratio, cutting down time churning these reports daily by 35%
- Led the product analytic team of 3 to design A/B testing to optimise product features and development
- Built some static pages on the current website with HTML, CSS and JavaScript
- Performed sentiment analysis using logistic regression and NLTK to identify the customers' painpoints, hence improving the CSAT score by 43%
- **Skills**: A/B Testing · Experimental Design · Product Management · Python · SQL · NLP · Predictive Modelling · Web Development · Stakeholder Management · AWS

GOOGLE

Product Data Scientist, Intern *Jan 2020 – Apr 2020*

- Developed AB testing to analyze the google ad performance
- Built production level **k-means clustering** model for the marketing team to identify users for targeted advertising.
- Directly involved in end-to-end ML pipeline and got exposed to Google Cloud Platform

NUS INSTITUTE OF DATA SCIENCE

Data Scientist, Intern *Aug 2019 – Dec 2019*

- Deployed **Long Short Term Memory (LSTM)** algorithm to improve the translation of text
- Some other ad-hoc **text-mining** tasks involving deploying NLP to analyse news articles
- Analysed survey/questionnaire data to evaluate the effectiveness of new teaching techniques on student performance.
- **Handled missing data** and extract insights from datasets consisting of both numerical and categorical data; conduct **hypothesis testing** for significance
- Conducted data workshops on introductory Python, R, Machine Learning for undergraduate students
- ***Technology Stack***: Python, R, ggplot, matplotlib, pandas, NLP

PERROMART | E-commerce

Data Analyst Intern *May 2019 – Aug 2019*

- Worked with Customer Service team to implement NPS as a performance metric
- Created a 3-tier loyalty programme based on revenue, LTV of customers, leading to a 5% increase in sales and 2-3% increase in retention rate.
- Performed *Cohort Analysis* on Excel to understand customers lifecycle, retention rates and behaviour for actionable insights
- Used Python and R to extract and analyse data and communicated actionable insights to the marketing and product development teams
- Wrote Macros to **automate tasks** for non-technical teams, increasing efficiency and productivity
- Implemented Prophet Time Series model in **R** to forecast the sales for 11/11 event
- ***Technology Stack***: Python, R, Excel/Macros, Tableau

EDUCATION

National University of Singapore (NUS) *Jul 2016 – Apr 2020*

- Bachelor of Science (Honours) in Statistics
- Relevant courses: Probability, Machine Learning, Categorical Data Analysis, Bayesian Statistics, Monte Carlo simulation, Experimental Design.

Data Science Bootcamps

- Deep Learning Specialization: Deep Learning.AI x Coursera
- NLP Specialization: Microsoft x EdX
- NLP Specialization: Deep Learning.AI x Coursera

Full-stack Development Bootcamp

- Relevant concepts: HTML, CSS, Javascript, React, MongoDB, Express, Node.js, Rest APIs

DATA SCIENCE PROJECTS

Sentiment Analysis of customer reviews

- Scraped the data from google reviews with Scrapy and Python
- Used NTLK and logistic regression to classify positive and negative feedback, deployed the model to production using AWS Sagemaker
- Presented the findings to key stakeholders, including BOD and investors

Recommendation system (for our agents to recommend to customers)

- Used Cosine similarity as a metric to find the most relevant products to recommend to the customers.
- Built the front-end with Flask (Python) and deployed on Heroku for internal usage

Recommendation system (for customers to use) [in-progress]

- Use collaborative filtering and matrix factorisation to find relevant products based on other customers interaction towards the product
- Building a pipeline that handles the streaming of new customer data to improve recommendation system

Customer Churn prediction

- Model used: random forest
- Improved the model accuracy from 83% to 89% through feature engineering and regularisation techniques

Text Translation (personal project)

- Translate text from English into Vietnamese and vice versa using Seq2seq network
- Codes documented in Jupyter notebook only, no deployment to production

Difference-in-difference modelling

- Modelled the impact of a campaign/product feature on the conversion of the customer using statistical DID modelling