Yiming Zhang

(+86) 187-0152-1223 | yimingz8@student.unimelb.edu.au
www.yimingzhang.me | Github: yiming95

EDUCATION

The University of Melbourne *Feb 2018 - Dec 2019*
Master of Information Technology *Melbourne, Australia*

The University of Nottingham *Sep 2014 - Jul 2017*
BSc. Hons Computer Science with Artificial Intelligence, First Class Honours *Nottingham, UK*

The University of Nottingham Ningbo China *Sep 2013 - Jul 2017*
BSc. Hons Computer Science with Artificial Intelligence, First Class Honours *Ningbo, China*

RESEARCH INTEREST

- Machine learning/ Deep learning
- Big Data
- Bioinformatics

RESEARCH PROJECT EXPERIENCE

Master Thesis: Machine Learning for Fertility Prediction *Jun 2019*
Machine Learning, Data Mining *Melbourne, Australia*

- Generated a model for predicting the probability of fertility after a patient has received certain cancer treatment.
- Experimented six machine learning algorithms K-Nearest Neighbour (KNN), Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF) and Artificial Neural Network (ANN) to predict the fertility status of the patient.
- Used methods such as Chi-squared based feature selection, 10-fold cross-validation, and learning curve to process the data and analyze the models.
- The best performing model was KNN with the highest accuracy of 0.7661, highest precision of 0.7752, highest recall of 0.7662, and highest F1 score of 0.7557.

Automatic Fact Verification *May 2019*
CodaLab Competition, Information Retrieval, NLP, Deep Learning *Melbourne, Australia*

- The Automatic Fact Verification system is a system that can automatically validate whether a claim is true, false or unverifiable based on the information in a large text corpus
- The system can find the relevant sentences which are related to the query in the corpus and then use the evidence to predict the validity of the claim
- Applied a Rule-Based approach to retrieve the documents, then selected the evidence from the sentences which have top 7 TF-IDF similarity and filter them with a fine-tuned BERT model.
- Predicted the labels by using another fine-tuned BERT model, and the final performance is Document Selection F1 80.9%, Sentence Selection F1 72.0%, Label Accuracy 64.40%.

Big Data Analysis and Visualisation *Apr 2019*
IaaS, Spark, Sentiment Analysis, Data Visualisation *Melbourne, Australia*

- Explored "Wrath" by harvesting tweets from the Great Melbourne Area and then apply data analysis

techniques and machine learning algorithms to the data.

- Referenced and parallelized the tweets stored in CouchDB to create Resilient Distributed Datasets (RDD).
- Used multiple data visualization tools such as choropleth map, pie chart, bar chart to demonstrate the emotion and sentiment in each region of Greater Melbourne, and used a word cloud to find out the possible reasons of wrath.
- Trained a Multiple Linear Regression Model that allow us the make predictions of the tweet wrath rate for a region given its AURIN data.

Approximate String Matching *Oct 2018*

NLP, String Matching Algorithm *Melbourne, Australia*

- Analyzed a Wikipedia dataset and then identify the causes of the typographical errors of Wikipedia editors.
- Used six approximate string matching methods including edit distance algorithm, N-gram algorithm and phonetic algorithm to solve the typographical errors and spelling correction.
- Evaluated these algorithms on the dataset and the Soundex algorithm has the best recall which is 82.15%.
- The results indicated that two causes of typographical errors made by wiki editors are: substitute a letter for a letter adjacent on the keyboard and phonetic attempts at spelling or guessing.

PROFESSIONAL EXPERIENCE

DarkSpede PTY LTD *Jul 2019 - Present*

Web System Developer Intern *Melbourne, Australia*

- Participated (as a core developer) in the development of a social media mobile application.
- Optimized the product by using AWS Lambda function and AWS DynamoDB.
- Ran functionality testing and debug code.

Qunar.com *Sep 2017 - Dec 2017*

Data Analyst and Product Manager Intern *Beijing, China*

- Analyzed raw data from the Qunar backend system with Excel to help improving user experience.
- Proposing, experimenting and modeling hypotheses about market prediction.
- Participated in product design.

Languages

Name	Proficiency
Chinese	Native Proficiency
English	Full Professional Proficiency

MISCELLANEOUS

- **Primary language:** Python, Java, JavaScript
- **Knowledge of:** Pandas, Scikit-learn, TensorFlow, Keras, Linux