

# Yuxuan He(Jennifer)

444 Washington Blvd, Jersey City, NJ, 07310

Email : yuxuan.he.1@gmail.com

Mobile : +1-201-240-6397

## EDUCATION

---

- **New York University** New York, NY
  - **Master of Science in Data Science** **December 2020**  
Natural Language Processing Track, Data Science and Software Service Award
- **University of Richmond** Richmond, VA
  - **Bachelor of Science in Mathematics and Computer Science** **Jan 2018**

## WORK EXPERIENCE

---

- **Naspers Limited - NYU** New York, NY
  - **Data Scientist Intern** **Sep 2019 - Feb 2020**
    - Build FinBERT, a fine-tuned language model based on BERT specialized in financial domain
    - Built a multi-task language learner based on FinBERT to solve multiple NLP tasks in finance domain using Model-Agnostic Meta-Learning(MAML) method using **PyTorch**
    - Fine-tune language learner for Naspers' business and financial tasks and improve the benchmark performance on pre-defined set of tasks
- **NYU Center of Data Science** New York, NYU
  - **Research Assistant** **April 2019 - Feb 2020**
    - Implemented FairBench, a machine learning framework which supports data ingestion, custom feature transformation, data cleaning and reusing of fairness metrics and fairness enhancing strategies using **Python, ScikitLearn, IBM AIF360**
    - Conducted extensive experiments on state-of-art machine learning methods and fairness-enhancing methods on multiple datasets and summarized finding about resulting metrics measures in paper
- **Linkedin Corporation** Sunnyvale, CA
  - **Data Engineer Intern** **Jun 2016 - Sep 2016**
    - Designed an internal data logging tool that supported **securely logging, indexing, searching, storing, visualizing and analyzing** of billions of confidential data records distributed stored across multiple data centers using **Java**
    - Implemented the data logging tool and successfully reduced teams average time to detect error using **Couchbase, Kafka, Elasticsearch, Restli filters and Kibana dashboard**

## PROJECT EXPERIENCES

---

- **Emotional Chatting Machine** [\[Github Repo\]](#)
  - Incorporated emotion to **Seq2seq** chatbot model by 1) trained sentiment classifiers using **BERT and Bi-LSTM** and selected the most accurate one to pre-classify corpus emotion using **TensorFlow**, 2) add memory module to decoder Gated RNN to handle emotion input using **Pytorch**
  - Further improved model evaluation scores by implementing global attention mechanism and deployed the model online using **Django**
- **Kaggle Million Song Dataset Challenge** [\[Github Repo\]](#)
  - Built a **Recommender System** using **alternating least squares (ALS)** model and ranked **10%** for Kaggle Million Song Dataset using **Python, Spark, SQLite**
  - Implemented cold-start strategy to handle new data occurrences
- **Restaurant Inspection Forecast:** [\[Github Repo\]](#) Conducted full data science pipeline 1) Crawled restaurant reviews using **Yelp and Goolge Map API** , 2) Extract features from geological, time and syntax data 3) Built and selected models to forecast restaurant inspection outcomes 4) Produe final evaluation and report(final acc 0.78%) using **Python, ScikitLearn, LightGBM, XGBoost and Tableau**

## SKILLS

---

- **Math and Statistical Skills:** Convex Optimization, Time Series Analysis, Statistics and Probability, Linear Algebra, A/B testing
- **Computer Skills:** Python(xgboost, tensorflow, pytorch, Keras,scikit-learn, numpy, pandas, opencv), Java, R, Hadoop, Spark, AWS, SQLite, SQL, ElasticSearch, Couchbase