

# Joshua Sun

2B Software Engineering Student

☎ (+1) 415-881-0388 | ✉ joshuasun1998@icloud.com | 🌐 jsun98 | 🏠 joshsun.ca

## QUALIFICATIONS

---

<b>Languages:</b>	Java, Python, JavaScript, C/C++
<b>Machine Learning:</b>	Vowpal Wabbit, TensorFlow, Scikit-learn, Gradient Boosting Decision Trees (GBDT)
<b>Big Data Analysis:</b>	Hadoop, MapReduce, HDFS, Hadoop-Streaming, Pig, Oozie, Storm, Avro
<b>Tools:</b>	Git, Maven, Docker, Travis CI, Screwdriver, Jenkins, Heroku
<b>Databases:</b>	MySQL, MongoDB, NoSQL

## EXPERIENCE

---

### Yahoo! Backend Software Engineer | Mail Anti-Spam Sunnyvale, California

The Anti-spam team analyzes **tens of billions** of SMTP requests daily for Yahoo!, AOL, and Oath.

Jan 2018 - Apr 2018

- Built a **terabyte-scale** distributed machine learning pipeline on Hadoop and wrote a series of MapReduce jobs in Java and Pig for data generation, feature selection, model training and model evaluation.
- Scaled Vowpal Wabbit and GBDT libraries to **parallel process 10 billion training examples** in under 30 minutes by using Hadoop-Streaming, MPI AllReduce and the spanning tree, reducing training latency **by 280%**.
- Enhanced the existing inbound spam detection system by adding a **Random Forest spam classifier**, resulting in an additional **150 million** spam messages captured daily and decreased the false-positive rate by 11%.
- Designed an Apache Storm duplicate detection bolt using Guava Cache for an online spam detection algorithm, increasing throughput of the cloud message queue by 65%.

### MentorMap Co-Founder [🔗](#)

Toronto, Ontario

MentorMap is a startup that connects prospective students with current university students.

May 2017 - Nov 2017

- Architected a **video scheduling web app** for mentors and mentees using Node.js, MongoDB, HTML and CSS.
- Designed a scalable MongoDB/Mongoose model that avoided transactions by using embedded documents.
- Led a team of 5 in software development and marketing and recruited **100+ members**.

### OMERS Full Stack Developer | Test Automation Team

Toronto, Ontario

OMERS is Canada's largest pension fund with \$72 billion in net assets.

May 2017 - Aug 2017

- Developed React/Redux/Node.js web applications to simplify the mock data creation workflow for 40 testers.
- Optimized the Selenium Grid infrastructure to efficiently run tests in parallel, reducing runtime by 45%.

## PROJECTS

---

### Stock Predictor + Trade-bot Python, TensorFlow, Scikit-learn, MySQL [🔗](#)

- Applied supervised learning methods such as Neural Networks, Random Forest and Gradient Boosting regressors to forecast the S&P 500 stock prices based on past returns and financial indicators.
- Developed a simulation trade-bot that buys and sells stocks based on the model forecasts.

### Lacs/LISP Compilers Scala, C++ [🔗](#)

- Implemented the CYK parsing algorithm, heap/stack memory management and a garbage collector.

## EDUCATION

---

### University of Waterloo

Sep 2016 - Apr 2021

Bachelor of Software Engineering, Joint Combinatorics and Optimization

CGPA: 3.70

## HONORS

---

**6th Place @ YHack** Competed against 115 teams at the Yale University hackathon

Dec 2017

**Governor General's Academic Medal** Highest GPA in high school of 439 students

Jun 2016