

# DENIZ TURKCAPAR

San Francisco, CA ▪ [dturkcapar@gmail.com](mailto:dturkcapar@gmail.com) ▪ <https://denizturkcapar.github.io> ▪ +1 773 524 8459

## EDUCATION

---

### The University of Chicago

Bachelor of Science in Data Science Specialization in Economics and Computer Science

Expected, Jun 2021

**Cumulative GPA:** 3.5 / 4.0

**Honors & Accomplishments:** Dean's List 2017-2020, Facebook University Data Analytics Batch Day Finalist 2019, Google Engineering Practicum Finalist 2019, Grace Hopper Scholar 2019-2020, Jeff Metcalf Fellow 2017-2020

## SKILLS AND RELEVANT COURSEWORK

---

**Programming Languages / Skills:** Fluent in Python, SQL, R, Java; experienced in C, Ruby, JavaScript, C++, statistical modeling, A/B testing, market/user research, data mining, project management

**Technologies and Frameworks:** Apache Kafka, Docker, Kubernetes, AWS (EMR, S3 & Elasticsearch), Hadoop, Spark, Git, REST API

**Relevant Courses:** Machine Learning, Data Science for Computer Scientists, Big Data Analysis, Software Development, Theory of Algorithms, Discrete Math, Statistical Models/Methods, Econometrics, Linear Algebra, Calculus

## WORK EXPERIENCE

---

### Salesforce

San Francisco, CA

Software Engineer Intern in the Search Data Division

Jun 2020 - Aug 2020

- Developed Sort Rules feature end-to-end for Business to Business Commerce Search, which is one of the three major features to be rolled out in the next release, within the search indexing schema using Java, SQL, Elasticsearch and Spark, enabling sorting by alphabetical order for the first time
- Implemented RESTful APIs for sort rules retrieval and display in the UI, significantly decreasing the time it takes for the consumer to not only set up the webstore's sort rules, but also receive sorted products
- Wrote automation tests to cover any possible regression with the introduction of the new feature, committing to 100% no regression

### PayPal

Chicago, IL

Software Engineer Intern in the Data Engineering Division

Jun 2019 - Aug 2019

- Implemented a Cronjob in Kubernetes to send heartbeat messages every 5 seconds and defined a DataDog plug-in, enabling the team to have data on Kafka topic lags in an interactive graph in terms of seconds of lag rather than message count for the first time
- Created and deployed sane memory related defaults and a dynamically adjusting heap size for certain topics in Kafka to eliminate crash loop back off by over 95% in Kubernetes pods
- Maintained data streaming replication over multiple data centers and AWS regions to make all Kafka data (across physical data centers & AWS) readily accessible to all teams and applications

### Becker Friedman Institute for Research in Economics

Chicago, IL

Data Scientist

Sep 2018 - Jun 2019

- Performed data cleaning and data mining using association and outlier detection via Python, R, and SQL to investigate the influence of the perception of Saudi husbands on the female labor force participation in Saudi Arabia
- Implemented scalable algorithms for similarity detection across analysis units in Python (Pandas) with ~60% gain in speed
- Worked on feature engineering and analyzed several datasets in detail to detect anomalies, outliers, and underlying patterns to increase model performance with ~43% faster results and increased accuracy in detecting similarities

## PROJECTS

---

### Matching with Predicate Constraints – Data Science Researcher at CHIData Research Group

Feb – Jun 2020

- Conducted research on Fuzzy Matching Problems with Predicate Constraints, such as missing data, using Graph Theory to improve both the efficiency and the search outcome of matching in data science problems
- Implemented a bipartite matching algorithm that suggests search results, achieving 90% accuracy in search suggestions

### Modeling Opinions and the Female Labor Force Participation Gap in the U.S.

Mar - Jun 2020

- Evaluated which data features are relevant in modeling labor force participation and the labor force participation gap (using techniques such as LASSO);
- Implemented a kNN model that was 76.1% accurate in predicting female labor force participation using a variety of features

### Predicting Yelp Elite status

Mar - Jun 2019

- Ingested large amounts of Yelp data via MapReduce jobs and used natural language processing (NLP) to assess whether elevated language in reviews help to become an Elite Yelp member
- Trained a logistic regression model using various hypothesized key factors to predict Elite status with 97.8% prediction accuracy, and a k-nearest-neighbors algorithm to find the most similar users with 95% accuracy in grouping Elite Yelp members together

## LEADERSHIP ACTIVITIES

---

### Girls Who Code

San Francisco Bay Area, CA

Teaching Assistant

Jun 2018 - Aug 2018

- Taught a class of 20 minority female high school students, all of whom later placed into top colleges to study computer science, utilizing a project-based curriculum on foundational computer science at Walmart Labs to develop programming skills and soft skills