

# Kishan Thakkar

kishanthakkar28@gmail.com – 847-975-1783 – Schaumburg, IL – 1kishan.github.io

## EDUCATION

### UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

B.S. COMPUTER SCIENCE | DEAN'S LIST SPRING 2018 - 2020

AUG 2017 - DEC 2020

GPA: 3.76/4.00

Relevant Coursework:

Data Structures • Algorithms • Systems Programming • Computer Architecture • Artificial Intelligence  
Programming Languages & Compilers • Deep Learning • Parallel Programming • Distributed Systems

## SKILLS

**DEVELOPMENT:** JAVA, PYTHON, SCALA

**SYSTEM:** C, C++

**TOOLS:** GIT, JUNIT4, LOG4J2, AWS

**DEV OPS:** BAZEL, GRADLE, MAVEN, FLASK

**DATABASES:** NOSQL, SPARKSQL, REDIS

**DATA PROCESSING:** APACHE SPARK, NUMPY, PANDAS

## EXPERIENCE

### PINTEREST | SOFTWARE ENGINEERING INTERN

MAY 2020 – AUG 2020 | SAN FRANCISCO, CA

Scala, Python3, Spark, SparkSQL

- Built a data workflow, using a series of Apache Spark jobs, that would identify Pinterest pages crawled by the Googlebot, determine interlinks originating from these pages, and calculate aggregated metrics about the crawled and linked pages.
- Utilized the data workflow to collect one month's worth of crawl data . The workflow identified over 70 billion Pinterest pages crawled by the googlebot and over 200 billion interlinks originating from these crawled pages.
- Reduced the overall run time of the data workflow by 40% by eliminating data skew which allowed the Spark jobs to run with a greater degree of parallelism.
- Leveraged the Spark GraphX library to build a directed graph representing the Pinterest pages identified from the data workflow and the interlinks between them; Used GraphX to also run PageRank on this graph which assigned a score to each page in the input graph allowing us to rank the pages from "best" to "worst"
- Used inferential statistics and correlation analysis to identify relationships between the generated "ranking" of a page and the amount of traffic that page received.

### AMAZON | SOFTWARE DEVELOPMENT ENGINEER INTERN

MAY 2019 – AUG 2019 | SEATTLE, WA

Java, AWS, Redis, NoSQL

- Designed and bootstrapped a cloud-based in-memory database which served as a read-replica for a NoSQL database fundamental for AWS Batch job scheduling.
- Developed an "event applier worker" application which maintained eventual consistency between the two databases by propagating changes from the NoSQL database to the in-memory database.
- Used Java multi-threading to increase throughput of the worker by 50%; maintained a thread safe application by using appropriate design patterns and data structures.
- Engineered a monitoring protocol which would trigger the appropriate failure recovery mechanism depending on a complete or partial failure of the in-memory database.
- Iteratively developed modularized packages while writing appropriate tests with JUnit and logging using Log4j2.

## PROJECTS

### FASHION IMAGE CLASSIFICATION | PYTHON, NUMPY, DEEP LEARNING

APRIL 2019

- Implemented a fully connected 4-layer neural network and trained a model using Mini-batch Gradient Descent.
- Decreased training time by using Numpy functions to perform calculations within the neural-network.
- Achieved 92% accuracy in labeling input images to their correct value (hat, shirt, socks, etc ...)

### NLP STOCK PREDICTOR | PYTHON, PANDAS, SCIKIT

MAY 2019

- Performed sentiment analysis on Reddit comments about Fortune 500 companies using TextBlob.
- Built a regression model to find trends between the "polarity" (change in sentiment over the course of a week) for a company and its historical change in price during that time span.