Harsha Somisetty

Edison, NJ | 732-707-2001 | harsha.somisetty@rutgers.edu
harshasomisetty.me | linkedin.com/in/harshasomisetty | github.com/harshasomisetty

EDUCATION

Rutgers University - Honors College   **New Brunswick, NJ**
Majors: Computer Science, Math, Statistics   *September 2019 - December 2022*

- **GPA:** 3.79 (159 credits, 120 to graduate) **Relevant Coursework:**
- **Graduate:** Algorithms, Comp Networking, Machine Learning, Massive Data Mining, Massive Data Storage, Time Series Analysis, Design of Statistical Experiments
- **Undergrad**: Topological Data Analysis, Operating Sys, Stochastic Processes, Real analysis, Lin Alg (proofs)

RESEARCH

Undergrad Researcher - Topological Data Analysis (TDA) for Financial Time Series (FTS)   **Edison, NJ**
Lead Student Researcher with Prof. Shane Kepley   *May 2021 — Present*

- Explored TDA for market transitions and crashes, implemented code for 3 research papers in **Python**
- Created novel representation of orderflow (230 million trade datapts) to volume bars to track price & volume
- Constructed new TS signal by calculating L1 norms of return point clouds using **GiottoTDA** and **Pandas**
- Improved state-of-the-art latest methods with new signal reacting notably faster to extreme market events/crashes

PROJECTS

AutoBlog: Twitter Threads to Ready-made Blogs   **Edison, NJ**
Link to demo at twitblog.xyz   *Oct 2021 — Present*

- Created an algo to detect Twitter threads from millions of tweets, and used topic modeling to identify core topics
- Built frontend with **React/Tailwind**, backend with **NodeJS/Express**, deployed with **Docker/Nginx**
- Advertised on Twitter and got over 1000+ hits, working with content creators to build out more features

Checkra: Podcast Summarizer & Topic Timestamp Generator: **Edison, NJ**
Link to demo at checkra.xyz   *Sep 2020 — Present*

- Implemented TF-IDF summarization methods in **Python** on 100+ longform podcasts to identify key information
- Transcribed Youtube podcasts with **Pyspark**, auto-generated timestamps, extracted relevant keywords with **Spacy**
- Created a new Latent Dirichlet Allocation algo for topic modeling and finding podcasts subsections/timestamps
- Constructed various dashboards with **Plotly**, deployed website using **Flask**, stored data with **MongoDB**

Pagerank Stock & Cryptocurrency Trading Signal Generator   **Edison, NJ**
Solo Project   *Jan 2021 — Present*

- Created pipeline to get 1 mil+ stock Tweets, extracted asset tickers and sentiment with **Spacy** and **PySpark**
- Discovered reputable Twitter users based on historical accuracy, downloaded 300k follow relations between users
- Stored data in **Neo4j**, found stocks using Pagerank on reputation of author and frequency of stock mentions
- Paired Pagerank scores with trading indicators, winning manual trades 70% of the time (but during a bull market)

SKILLS & INTERESTS

- **Skills:** Python, R, C, Java, Javascript, Solidity, Typescript, Webscraping, LaTeX, Bash, Git, Emacs, Vim/TMux
- **Tools:** NodeJs, SQLite, Neo4j, MongoDb, Apache Spark, React, Flask, Pandas, Heroku, Spacy
- **Interests:** Calisthenics, Guitar, Geopolitics, History, Life Extension, Podcasts, Reading, Rubik's cubing