Kartik **Nautiyal**

Boston, MA

+1 774-502-1533 | knautiyal@wpi.edu | kartiknautiyal.github.io | kartiknautiyal | kartiknautiyal

Education

Worcester Polytechnic Institute *Worcester, MA*

Master of Science in Data Science (4/4) *Aug '21 - May '23*

- Database Management Systems, Machine Learning, Statistical Methods for Data Science, Big Data Management, Data Visualization

Mumbai University *Mumbai, India*

Bachelor of Engineering (8.58/10) *July '15 - Aug '19*

Work Experience

John Hancock Life Insurance Company *Boston, MA*

Advanced Analytics Data Science Co-op (Retirement Plan Services)(SQL Server, Python: Pandas, Scikit-learn, FastAPI) *Jan'23 - Present*

- Training an ML model to predict which participants will consolidate their old 401(k) accounts with the company
  – Increase the cash flow into the company. (Results awaited)
- Building an API end-point using FastAPI for the sales team to leverage the results of the model, enabling a secure & quicker data access
  – Allow faster use of results from the model thereby reducing lead time

U.S. Army DEVCOM at Worcester Polytechnic Institute *Worcester, MA*

Graduate Student Researcher (Python-FastAPI, Javascript, PostgreSQL) *Jan'22 - May '22*

- Worked with the material science unit of the DEVCOM to develop features for a web-based platform to make data access for material scientists easier
- Feature 1: Developed a feature to upload, validate, parse and put data from the front-end using a graphic interface to the database, if valid
  – Reduced the random error rate by 62%
- Feature 2: Developed another feature that involved fetching an image from the back-end given a data point at the front-end using a hover function
  – Material Scientists were able to perform the same jobs faster using the feature

Data Management and Analysis Vertical, NITI Aayog, Govt. of India *Delhi, India*

Data Science Intern (Python (TensorFlow), JavaScript) *Oct '19 - April '20*

- Researched and found a method of measuring relative GDP contribution (with a correlation of 70%) by using night-light raster data from Google Earth Engine and developed a web-tool using JavaScript to visualize it on map
  – Allowed researchers to study the impact of some of the schemes of Govt. of India without relying on census data
- Implemented Support Vector Machine algorithm to identify different terrains from satellite images with an accuracy of about 81%
  – Helped researchers estimate agricultural productivity of an area

Projects

Economic Hotspot Explorer (Web-Tool) *Link*

Research Project at NITI Aayog (JavaScript)

- Researched and found a proxy indicator for district level GDP in night-time lights which showed a 70% correlation
- Used night-time lights data to segregate cities in India into areas of low, medium and high economic activity
- Developed an interactive web-tool using JavaScript to visualise and create interactive digital maps using GIS data on Google Earth Engine

Covid Risk Tracker (Analysis) *Link*

Academic Project at WPI (Python: Pandas, Scikit)

- Studied the link of 27 Socio-Economic indicators and Covid-19 risk by analysing the first 6 months of Covid data (most important time to respond to pandemics according to epidemiologists) in all cities in the state of Massachusetts
- Categorized cities by risk using K-means clustering algorithm on Covid data and using the elbow method to select optimal value of K
- Trained multiple supervised learning algorithms to predict risk categories in communities using socio-economic information as predictors and achieved an accuracy of 69% for 4 risk categories and 90% in case of 2 risk categories

NYC Yellow Taxi Cab Ride Analysis (Real-time dashboard, ETL Pipeline) *Link*

Academic Project at WPI (Docker, Amazon S3, Spark structured streaming, Kafka, MongoDB, Mongo Atlas Charts)

- Created a real-time dashboard on Mongo Atlas Charts by implementing an ETL Pipeline which streams data from AWS S3 bucket using Kafka and then then analyses by Spark Structured Streaming which then populates a MongoDB database.
- All technologies were implemented inside docker containers which was connected by an orchestrator container containing Python
- The dataset was sourced from kaggle and replicated to around 18GB and the entire application was implemented on an 8GB machine.

Visualizing Consensus (Visualization Website) *Link*

Academic Project at WPI (JavaScript: D3.JS)

- Proposed a new method for consensus ranking visualization used for decision making using D3.JS using features like scrollytelling
- Increased understandability in the agreement/disagreement in a consensus with respect to individual rankings and allowed comparison among multiple consensus rankings

Skills

- Programming and Database: Python (Pandas, NumPy, Sci-Kit learn, TensorFlow, FastAPI), SQL (Oracle DB, MySQL, PostgreSQL), JavaScript, R, MATLAB
- Big Data: Hadoop, Spark, Kafka, MongoDB, Open MPI, Amazon S3
- Data Visualization: Microsoft Power BI, Python (Matplotlib, Seaborn), JavaScript (D3.JS), Google Earth Engine (Javascript)
- Tools: Git, Visual Studio Code, Docker, IBM SPSS Modeler, Microsoft Office (Excel, Word, Powerpoint), Jupyter Lab, Jupyter Notebooks
- Experience: Exploratory Data Analysis, Data Mining, Processing large Datasets, Classification, Regression, Clustering, Decision Trees, Text Analysis