Nischal Mahaveer Chand

Boston, MA | nischal.bhatewara@gmail.com | (857) 277-9145
Blog: medium.com/@darkestfloyd | LinkedIn: nischalmc | GitHub: DarkestFloyd

WORK EXPERIENCE

BERG LLC, Framingham, MA — Nov 2019 - present
Data Scientist II
Tools: R, RStudio, Shiny, tidyverse, git, linux, ganglia

- **Statistical data analysis** (regression analysis, survival analysis, hypothesis testing and multiple testing):
  - Perform survival analysis on cohort of n>100 pancreatic cancer patients to understand effectiveness of various chemotherapy and adjuvent treatments using Kaplan-Meier curve in R, resulting in a poster presentation at ESMO 2020.
- **Data pipelines** (ingest RESTful API (JSON), SQL databases, excel sheets, webscraping; orchestration using CRON):
  - Built automated data processing pipeline to process longitudinal clinical data of n>200k patients, resulting in discovery of various clinical features that impacted survival outcomes for COVID-19 patients by bAIcis®.
  - Built data processing pipeline to gather data from multiple data sources - clinical data using RESTful API, omics data from SQL servers and excel files - for n>450 patients, further performing QC checks and merging data based on timepoints.
  - Built pipelines to obtain external Patent data (uspto.gov), DrugBank data (drugbank.com), PubMed data (NCBI) and cross-reference with internal biomarkers for use by upper management, cutting down time required to find key markers.
  - Improve legacy SQL data pipeline by applying filters and aggregation in SQL to cut down processing time by 68%.
- **Data dashboards and visualizations** (R Shiny dashboards; ggplot2 visualizations):
  - Built interactive data dashboard to allow clinicians to examine impact of clinical features on survival outcomes for various subpopulation of COVID-19 patients (by age, race, comorbidities, etc), further deployed dashboard on Azure cloud.
  - Built R-Shiny dashboards to track key metrics and provide overview of internal processes and data stores.
  - Create data visualizations of analysis methods and results for publications and posters using ggplot2 in R.
- **Deep learning**: Trained over 2k deep neural network models to optimize running time and precision of bAIcis®using keras, cutting down on-boarding time of new team members getting started with bAIcis®.
- Provide support to Analytics team as Linux administrator for Analytics Server (Turing) and HPC cluster of 55 CPU nodes and 1 GPU node (Bayes cluster), further deploy and test new versions of bAIcis®on cluster.
- Contribute to company IP via software implementations and improvements to AI platform and legacy dashboards.

Marcus Institute for Aging Research, Hebrew SeniorLife, Roslindale, MA — July 2018 - Dec 2018
Junior Data Scientist / Co-op Student

- *AD Supplement*: - Performed regression analysis and hypothesis testing on clinical trials data using R and SQL.
  - Actively communicated results to directors and researchers.
- *Smartphone Lab*: - Built data processing pipelines for time-series sensor data using R and SQL.
  - Improved data processing speeds by upto 13%.
- *shinyMRI*: - Built scalable R shiny application to visualize 3D and 4D MRI images.
  - Improved diagnosis speed for in-house Alzheimer's research.
  - Application received honorable mention by RStudio, Inc.

EDUCATION

NORTHEASTERN UNIVERSITY, Boston, MA
Khoury College of Computer Sciences, GPA: 3.79/4.0 — Sept 2017 - Aug 2019
Master of Science in Data Science
Related Courses: Supervised Machine Learning; Unsupervised Machine Learning and Data Mining; Data Visualization; Natural Language Processing; Statistics for Bioinformatics; Algorithms
TA positions: Data Management in R (Spring 2019); Unsupervised Machine Learning (Summer 2019)

ALLIANCE UNIVERSITY, Bengaluru, India — 2013 - 2017
College of Engineering and Design, CGPA: 3.4/4.0 — Aug 2013 - June 2017
Bachelor of Technology in Computer Science and Engineering
Related Courses: Data Mining and Data Warehousing; Big Data Analytics; Design and Analysis of Algorithms

PUBLICATIONS

- (Co-author) Kiebish, M.A., Tekumalla, P. et al. Clinical utility of a serum biomarker panel in distinguishing prostate cancer from benign prostate hyperplasia. Sci Rep 11, 15052 (2021). https://doi.org/10.1038/s41598-021-94438-4
- (Co-author) Miller, Gregory M., et al. "Ondansetron use is associated with lower COVID-19 mortality in a Real-World Data network- based analysis." medRxiv (2021).
- (Co-author) Moser, A. J., et al. "1485P Identification of novel protein biomarkers for FOLFIRINOX-based chemotherapy response in advanced pancreatic adenocarcinoma using patient omics and Bayesian AI." Annals of Oncology 32 (2021): S1094.
- (Co-author) Moser, A. J., et al. "1547P Project Survival®: High-fidelity longitudinal phenotypic and multi-omic characterization of pancreatic ductal adenocarcinoma (PDAC) for biomarker discovery." Annals of Oncology 31 (2020): S947-S948.

RELEVANT PROJECTS

CardinalVis - Dynamic data visualization for Mass Spectrometry Imaging (MSI) experiments May 2019 - Aug 2019
Project Repository: https://github.com/kuwisdelu/CardinalVis/

- Created R shiny dashboard and modules in R.
- Application is open-sourced and intended to cut down research times in MSI experiment research.

MURA - Bone X-ray image classification and data visualization project Jan 2019 - Apr 2019
Website: http://ec2-3-212-216-62.compute-1.amazonaws.com/

- Trained and tuned several classification models using sklearn and PyTorch in Python.
- Obtained best models for each classified using Cohen's Kappa metric.
- Created RESTful API service to provide real-time classification and dynamic visualizations using Flask-RESTful in Python.
- Deployed application, models, and visualization to AWS EC2 instance.

NL2code - Natural Language to Python code generator Jan 2018 - Apr 2018
Project Repository: https://github.com/darkestfloyd/NL2code

- Scraped open-source Python repositories to gather training data using beautifulsoup in Python.
- Trained custom neural translation system using Theano in Python.
- Improved accuracy and BLUE score over base paper.

Flashlight - Real Estate Property Assessment Visualization for the City of Boston Oct 2017 - Dec 2017

- Created dashboard to visualize various aspects of real estate assessment values using R "Shiny" and Leaflet in R.
- Geocoded missing coordinate values using Open Address dataset and Google Maps Geocode API in Python.
- Created Python scripts to automate conversion of Zillow's shapefiles to GeoJSON format and gathering of meta-data.

Movie Recommendation System Apr 2017 - June 2017

- Implemented a Recurrent Neural Network using TensorFlow in Python to recommend movies to a user. By analysing the sequence of movies reviewed, we were able to achieve high scores in peer review.

HONORS AND ACHIEVEMENTS

- Panelist for Northeastern University UNCOVER COVID datathon. Sep 2020
- Presented *MURA visualization tool* at The Northeastern University Visualization Consortium. Apr 2019
- Honorable mention for *shinyMRI* by RStudio. Apr 2019
- Graduate speaker for TechTalks at Grad TechDay, Khoury College. Mar 2019
- Coordinated and organized CodeWars, a 24-hour hackathon as part of university fest at Alliance University. Oct 2015