# Nischal Mahaveer Chand

Boston, MA | nischal.bhatewara@gmail.com | LinkedIn: nischalmc
Blog: medium.com/@darkestfloyd | Google Scholar: Nischal Mahaveer Chand | GitHub: DarkestFloyd

## SUMMARY

- Senior Data Scientist with experince applying machine learning and deep learning techniques to biomedical and multi-omics data in durg discovery, theraputics, and diagnostics.
- Skilled in data-wrangling, machine learning, and statistical analysis techniques to extract value from messy datasets.
- Strong understanding of data visualization techniques and story telling to deliver key results to diverse audiences.
- Experienced in communicating with executive team to understand business needs and commercialization objectives to provide analytics solutions that help drive business decisions.
- Proficient in R, Python, SQL, Java, and linux based operating systems.
- Committed to improving patient journeys and outcomes as they battle disease.

## WORK EXPERIENCE

**BPGbio, Inc.**, Framingham, MA
Datasets: high-throughput multi-omics, NGS, drug sensitivity, EMR/EHR, molecular assay, public domain
Tools: R, Python, R Shiny, tidyverse, git, linux, tensorflow, keras, pandas, ggplot2, caret, postman, Azure, Docker

*Senior Data Scientist*                                                                                          Jan 2023 - present

- Researched and developed deep learning algorithms to provide computational support in compound discovery and indication selection pipelines by leveraging cell line omics and compound structural data in Python using Keras.
- Developed machine learning and visualization pipeline to support commercialization objective of diagnostic test with an emphasis on model interpretability in R (caret) and Python (sklearn).
- Created code book with best practices for utilizing machine learning in diagnostics for clinical use in MS Word.
- Prepared and presented key results and progress updates at company wide scientific meetings using MS PowerPoint.

*Data Scientist I/II/III*                                                                                        Nov 2019 - Dec 2023

- Developed data pipelines to process and integrate large datasets from disparate sources to support company decision making and guide diagnostics and therapeutics strategies in R (tidyverse) and Python (Pandas).
- Implemented and deployed deep learning solution to provide optimal parameters for running internal platform bAIcis to reduce operational costs and increase user productivity in Python using sklearn and Keras.
- Deployed data pipelines, models, and dashboards on Azure using Docker to enable external collaborators to access and use the pipelines for data-driven decision making.
- Performed statistical data analysis on longitudinal multi-omics datasets to reveal biologically relevant insights in R using tidyverse resulting in development of validation molecular assays, project funding, and publications.
- Scaled legacy data pipelines by identifying and incorporating multi-processing techniques to improve data processing speeds by upto 70% in R using parallel.
- Collaborated with internal R&D team to guide experimental design and provide statistical and computational support.
- Developed and deployed dashboards to provide automation capabilities for routine tasks for R&D teams and provide overview of key metrics for program managers to help guide project decisions in R uning RShiny.
- Communicated insights to diverse audiences including data scientists, clinicians, senior leadership teams, scientific teams, and external collaborators via presentations and reports using MS PowerPoint and MS Word.
- Developed, documented, and maintained internal packages and linux shell scripts to fill Analytics team needs.

*Interim Linux System Administrator*                                                                            Mar 2021-Nov 2022
Tools: ganglia, cron, ssh, linux, mpi, slurm, fstab

- Provided support to analytics team by managing software stack of internal production servers and onsite HPC cluster.
- Deployed internal bAIcis platform on Azure resulting in increased scale while staying compliant, enabling team to execute on more than one projects.

**Marcus Institute for Aging Research**, Hebrew SeniorLife, Roslindale, MA
*Junior Data Scientist / Co-op Student*                                                                          July 2018 - Dec 2018
Datasets: patient forms and cognitive test scores, MRI and fMRI, radiologist annotations, accelerometer senor time-series
Tools: R, SQL, tidyverse, git

- *AD Supplement*:     - Performed regression analysis and hypothesis testing on in-house patient data using R and SQL.
                                   - Prepared reports and actively communicated results to researchers and institute director.
- *Smartphone Lab*:   - Built data processing pipelines to clean and process time-series sensor data from wearable tracking device using R and SQL.
- *shinyMRI*:              - Built 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.

## PUBLICATIONS

- (Co-author) Miller, G.M., et al. Hypothesis-Agnostic Network-Based Analysis of Real-World Data Suggests Ondansetron is Associated with Lower COVID-19 Any Cause Mortality. Drugs - Real World Outcomes 9, 359–375 (2022).
- (Co-author) Kiebish, M.A., Tekumalla, P., Ravipaty, S. et al. Clinical utility of a serum biomarker panel in distinguishing prostate cancer from benign prostate hyperplasia. Sci Rep 11, 15052 (2021).
- (Co-author) Grund, Eric M., et al. "Project Survival®: Discovery of a molecular-clinical phenome biomarker panel to detect pancreatic ductal adenocarcinoma among at risk populations using high-fidelity longitudinal phenotypic and multi-omic analysis." Cancer Research 82.12_Supplement (2022): 5145-5145.
- (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).
- (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).

## 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

Achievements: - Judge and panelist for Northeastern University UNCOVER COVID datathon (Sep 2020)
- Presenter at Northeastern University Visualization Consortium (April 2019)
- Selected speaker for Northeastern University GradTech Day tech talk (March 2019)

TA positions: Data Management in R (Spring 2019); Unsupervised Machine Learning (Summer 2019)


**ALLIANCE UNIVERSITY**, Bengaluru, India

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

## RELEVANT ACADEMIC PROJECTS

**CardinalVis** - Dynamic data visualization for Mass Spectrometry Imaging (MSI) experiments          May 2019 - Aug 2019

*Project Report: https://github.com/darkestfloyd/CardinalVis_AppNote*

- Created R shiny dashboard and modules in R to visualize mass-spectrometry data.
- 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

Summary: End of course project to improve interpretability in black box ML models; worked in a team of 4.

Project Repository: https://github.com/darkestfloyd/DS5500Project

Tools: Python, PyTorch, torchvision, scikit-learn, AWS

- Implemented image processing pipeline for adding image rotation, scaling, etc to enlarge training set using torchvision.
- Trained and tuned deep learning classification models for prediction of bone condition using PyTorch.
- Integrated GradCAM technique to visualize localization map of gradients for model interpretation in PyTorch.
- Depluyed model and provide real-time prediction for user uploaded image on AWS.


**NL2code** - Natural Language to Python code generator          Jan 2018 - Apr 2018

Summary: End of course project to implement NLP techniques; worked in a team of 3.

Project Repository: https://github.com/darkestfloyd/NL2code

Tools: Python, TensorFlow, Theano, beautifulsoup, NLTK, git

- Implemented data ingestion pipeline to scrape Python code and associated comments from public GitHub repositories to gather training data in Python using beautifulsoup.
- Trained encoder-decoder models in Theano to convert english comments (natural language) to Python code.
- Improved accuracy and BLUE score over base paper by incorporating grammatical strucutre into encoder.