Pushkar Inamdar

# Professional Summary

Applied Scientist with proven track record in Statistics, Spatial Analytics, Big Data Analytics and Machine Learning including Deep Learning technology.

# Education

**Ph.D.**: Spatial Sciences (Earth and Atmospheric Sc.), 12/2017

**Mississippi State University** - Starkville, MS **Master of Science**: Geoinformatics, 05/2012 **Symbiosis International Institute** - Pune, India

# Work History

**Data Scientist**, 04/2018 - Current

**UCSF** – San Francisco, CA

* Worked simultaneously on multiple projects to implement data pipelines and machine learning models.
* Applied Agile methodologies to enhance collaboration and drive iterative development in data science projects, delivering impactful solutions tailored to organizational objectives.
* Worked with stakeholders to develop quarterly roadmaps based on impact, effort, and test coordinations.
* Applied statistical and algebraic techniques to interpret key points from gathered data.
* Coached and mentored data scientists on spatial statistics and data mining techniques.
* Implemented randomized sampling techniques for optimized surveys.
* Developed climate variables, including heatwaves, maximum temperature, surface reflectance, and forest fires, to analyze and assess impacts of climate change on public health.
* Automated process of street view audits with Neural Network models to detect and segment features in satellite images to develop advanced socioeconomic neighborhood database of San Francisco Bay region.
* Developed scalable classification and regression models leveraging machine learning and deep learning with Apache Spark. Parallelism of large 'Duns & Bradstreet' business data helped understanding relation between liquor business and Liver disease rate in the neighborhoods.
* Built text classification model using machine learning models such as Logistic regression, SVM, KNN, Random Forest, Ensemble methods and deep learning methods like CNN, RNN, LSTM on text dataset for analysis.
* Used cross-validation to test model with different batches of data, tuned parameters to find best parameters for model and optimized, which eventually boosted performance.
* Performed Survival/Risk analysis on time dependent datasets using Cox Proportional Hazards, Logistic regression, and Weibull regression models.
* Developed greenspace, blue space, and light at night measures to characterize neighborhoods using satellite images.
* Performed experimental analysis using frequentist and Bayesian hypothesis techniques to report effects of spatial clusters on cohort participants.
* Implemented space–time cluster analysis on geocoded cohort and EHR dataset to understand impacts of environmental exposures on human health.
* Applied Natural Language Processing (NLP) techniques to analyze text patterns from registry datasets.
* Performed data wrangling, string pattern analysis of electronic health records and cohort datasets.
* Created geospatial maps and performed complex statistical data analysis, including multivariate statistical analysis, multilevel modeling, dimensionality reductions, time series modeling, spatial autocorrelation.
* Set up SQL database on servers to store neighborhood data for query analysis.
* Discovered stories told by data to present information to scientists and health professionals.
* Developed intricate algorithms based on deep-dive statistical analysis and predictive data modeling.
* Applied loss functions and variance explanation techniques to compare performance metrics.
* Assessed accuracy and effectiveness of new and existing data sources and data analysis techniques.

**Research Assistant**, 08/2013 - 12/2017

**Mississippi State University** – Starkville, MS

* Conducted time series modeling on climate data, LIDAR, and SAR (radar data) to comprehend spatiotemporal patterns. Study revealed significant increase in rate of glacial mass change in South America.
* Processed and analyzed health professional datasets using SQL querying and analytics tools.
* Geographically visualized and reported clusters of increasing shortages of health professionals in Mississippi.
* Designed and instructed courses on Statistics, Spatial Mapping, Photogrammetry, and Remote Sensing Image Analysis.

**Data Analyst**, 12/2011 - 07/2013

**Primove Infrastructure Pvt Ltd** – Pune, India

* Database design and implementation experience for sustainable development of villages in MH, India
* Processed data from analog and digital resources to perform spatial analysis.
* Supervised and Unsupervised image classification to extract land covers and created interactive web maps for region of interest.
* **Programming:** Python, R, HTML, Scala, JavaScript

# Skills

* + **Database Management:** MySQL, MongoDB, PostgreSQL, MS Access
* **Machine Learning:** Regression, Random Forest, SVM, Boosting, KNN, K-means, CNN, ANN
* **Time Series Forecasting:** ARIMA, NN, Multivariate Forecasting with LSTMs
* **Probability and Statistical analysis:** Bayesian hierarchical models, frequentist, Space-time analysis, Floating catchment area model, Spatial Autocorrelation
* **Scalable Data Processing:** Amazon Web Services (AWS), Spark, HIVE, Google Vertex AI
* **Geospatial tools and Programs:** QGIS, ArcGIS, Google Earth Engine, GDAL, geopandas, rasterio, Fiona, Shapely, Cartopy, xarray, Leaflet, OpenCV
* **Other Programs:** SAS, SPSS, STATA, Tableau

# Research Publications

* Spatial clusters of extended-spectrum beta-lactamase-producing *Escherichia coli* causing community- onset bacteriuria due to repeat infections: cluster analysis from a large urban medical center, San Francisco, 2014–2020. *Antimicrob Resist Infect Control* **12**, 115 (2023).
* Role of neighborhood context in ovarian cancer survival disparities: current research and future directions.

*American journal of obstetrics and gynecology*, *229*(4), p.366 (2023).

* School racial segregation and the health of Black children. *Pediatrics*, *149*(5) (2022).
* Joint associations of race, ethnicity, and socioeconomic status with mortality in the multiethnic cohort study. *JAMA Network Open*, *5*(4), pp. e226370-e226370 (2022).
* Traffic-related air pollution and lung cancer incidence: the California Multiethnic Cohort Study. *American Journal of Respiratory and Critical Care Medicine*, *206*(8), pp.1008-1018 (2022).
* Association between airport-related ultrafine particles and risk of malignant brain cancer: a multiethnic cohort study. *Cancer Research*, *81*(16), pp.4360-4369 (2021).
* A Spectral Analysis of Snow in Mt. Rainier. *Journal of Geography and Geology*, *10*(3), pp.1-20 (2018).
* *A Reconciled Estimation of the State of Cryospheric Components in the Southern Andes and California Using Geospatial Techniques*. Mississippi State University (2017).
* Assessment of irrigation and agriculture potential of the Krishna River basin using geospatial techniques.

*Indian Journal of Science and Technology*, *9*(44), pp.1-9 (2016).

* Spatial Patterns of Glacier Mass Change in the Southern Andes. *Photogrammetric Engineering & Remote Sensing*, *82*(10), pp.811-818 (2016).