

Devika Nair

Reston, VA | 240 246 6804 | devika.t.nair@gmail.com | <https://devikanair.netlify.com/>

Summary

I am a recent graduate of the MS Data Analytics program within American University's Kogod School of Business. After completing my BS in Neuroscience at Franklin & Marshall College, I worked at Genetic Alliance, a nonprofit dedicated to advocacy on behalf of communities affected by genetic conditions. I joined the healthcare world, working for The Advisory Board Company, as a business analyst for the company's internal Salesforce development team. During the pursuit of my graduate degree, I pivoted from healthcare IT into local government, serving on Arlington County's new Process & Data Analysis team. After a year there, I now apply my data science skills through Virginia Tech's Data Science for Public Good program. ***My objective is to continue to conduct data science research on important public good challenges.***

Interests

Data Analytics, Data Science, Public Good, Public Health, Predictive Models, Statistical Analysis, Writing and Blogging, Public Speaking, Neuroscience, French Language

Education

Virginia Tech, Data Science for Public Good, 2018

American University, Kogod School of Business - MS Analytics (GPA 3.7), 2018

Franklin & Marshall College - BA Neuroscience (GPA 3.2), 2012

Work Experience

Data Science for the Public Good (DSPG) Program Graduate Fellow, May - September 2018

I served as a graduate fellow for Virginia Tech's Data Science for Public Good program, managing a diverse group of undergraduates from multiple disciplines (engineering, data analytics, social sciences). Together, we conducted research on vertically integrated teams (undergraduates, graduate fellows, post-docs, researchers, and sponsor) to discover, integrate, and analyze multiple data sources to address sponsor-driven public good questions. I worked with the students to showcase our methods and findings in posters and presented findings to sponsors regularly and at the DSPG Symposium in August 2018. My projects and the skills I learned are described below:

Building composite metrics to measure student performance at Virginia Tech.

The Provost's office asked our team to explore and identify trends to understand the university's ability to attract and retain students, especially those from diverse backgrounds. Specifically, the university wanted to better understand patterns in the enrollment decisions of admitted VT applicants and patterns in the degree completion behaviors of current VT students.

Worked with the students to discover, ingest, clean, and analyze applicant and student data to model these outcomes, using data from a wide range of sources, such as the National Student Clearinghouse, Virginia Department of Education, State Council of Higher Education of Virginia, College Board, US News, and Virginia Tech's University Data Commons (the university's own data lake housing operational and student data).

Led team to create and validate regression models to identify enrollment decisions and successful student progression behaviors. Implemented LASSO to reduce dimensionality and identify

meaningful independent variables across our integrated dataset. Applied a binomial logistic regression model to identify and characterize the relationships between our responses and the independent variables. Presented the results using novel visualization techniques. Based on the findings, we provided actionable recommendations to the Provost's office.

Discovering and ingesting new geographically-coded data to inform Virginia Cooperative Extension policy decisions.

The Virginia Cooperative Extension (VCE) requested visibility into sub-county data to address local challenges. I worked with the students to discover, evaluate, ingest, and translate useful federal datasets to inform VCE decisions around program investments and resource allocation.

Leveraged federally published datasets that provide information about natural resources: the Environmental Protection Agency's Air Quality Index and Radon data; US Geological Service's soil composition data; Mine Safety and Health Administration data; and US Department of Agriculture crop coverage data.

Worked with the team to calculate centroid locations for each census block group in Virginia. Used Inverse Distance Weighting (IDW), a data interpolation method in which unknown values at specific locations are approximated through weighting values at known locations. Computed productivity measures of soil fertility.

Mapped locations and productivity measures at the census block level and provided the data in a dashboard for easy access by VCE agents. Through our research, we demonstrated that we are able to provide a more dimensional and useful view of Virginia to assist VCE professionals in their work.

Exploratory Statistical Analysis of Postsecondary-Going Culture at Virginia High Schools.

Identified several federal, state, and local data sources to learn about the characteristics of high schools that encourage students to pursue postsecondary educational opportunities. These data sources include the Integrated Postsecondary Education Data System (IPEDS); Adult Training and Education Survey (ATES); Harvard and MIT MOOC providers; Virginia Department of Education (VDOE); State Council of Higher Education for Virginia (SCHEV); College Board; and Current Population Survey (CPS) data on Certificates and Licenses.

Examined hundreds of variables over several years of data. Assisted with using Principle Components Analysis (PCA) to select variables for use in predictive models. Used school characteristics, such as mean SAT scores and on-time graduation rates, to predict the proportion of students that will enter various postsecondary trajectories such as 4-year college enrollment, employment, and the military and mapped the data and model results to present visually.

Arlington County Government, Management Intern, Process and Data Analysis, April 2017 – May 2018

I worked as a full-time Management Intern for Arlington County's new Process & Data Analysis team, serving as a junior data analyst along with a Senior Data Analyst, another Management Intern (who served as a junior business analyst), and our team lead. Our team functioned to bring about a data vision for the county and I worked both independently and collaboratively to fulfill this purpose.

Created a number of data visualizations using Power BI reports and dashboards to meet the business goals of diverse teams across Arlington County Government. These included Department of Technology Services (device management report and Microsoft O365 utilization dashboard); Department of Management and Finance (procurement request pipeline report); and Department of Community Planning, Housing, and Development (resident neighborhood request report). Similarly, translated and updated the Community Report Card, given by the Arlington Partnership

for Children, Youth, and Families, from its original PDF format to a Power BI dashboard with the goal of storing longitudinal data in a single repository as well as be able to provide ongoing, online public access to the report.

Built and manipulated SQL queries in SSMS to clean and integrate datasets, including an analysis of real estate property value appeal outcomes showing the disparity of outcomes based on differential appeal avenues and their impacts on county revenue.

Developed process and assembled requirements for an intake and scoring approach for new project requests to streamline Process & Data Analysis team's operations and track long-term impact of team efforts.

Established the Arlington Data Analyst Community, an internal group for Arlington County data analysts to gather, share resources, learn tools, and network to further common goals. Facilitated and served as teaching aide for an all-day, on-site Dashboard in a Day training session in Power BI to this group.

Led a data analytics themed session of Camp HERricane, an educational workshop for young women leaders in high school to learn about emergency management. Partnered with the Department of Public Safety Communications and Emergency Management to develop materials and coursework.

Assisted with the build of data products by gathering requirements and conducting user testing. One was for Department of Human Resources that emailed HR liaisons timesheet errors to ensure payroll completion and compliance. Another was for the Department of Environmental Services that automated the collection and storage of waste disposal data to show trajectory of Arlington County Government's sustainability efforts.

DC United, Data Science Student Consultant, Washington DC, August 2017 – May 2018

DC United asked our team to investigate the link between game conditions and game outcomes; specifically, how stadium, weather, and travel impacted wins and losses. I worked with a team of fellow graduate students with American University over the course of 8 months to gather, clean, link, model, and visualize data to answer this question.

Utilized MLS match-level and player-level data from OPTA, a sports analytics company. Discovered and manually acquired weather and stadium data from the web.

Selected, cleaned, and transformed data using Excel, SQL, and R. Analyzed datasets for redundancy and created summary variables where necessary. Imputed travel data using distance calculations.

Implemented machine learning techniques to model outcomes relative to conditions. Implemented LASSO, Ridge, XGBoost to select for significant variables. Applied multiple linear regression to build a model predicting magnitude and directionality of game outcome.

Built a Shiny application using regression model to generate potential game outcomes based on potential schedule and weather possibilities. Also mounted visuals to demonstrate trends.

Presented findings to Major League Soccer executives interested in optimizing seasonal MLS game schedules for maximum game entertainment value.

The Advisory Board Company, 2014-2016, Washington DC

The Advisory Board Company was a healthcare and higher education consulting firm, with both a traditional research arm and an innovative technology side. I was hired in 2014 as a coordinator for the client-facing Crimson performance technologies division but rose in the ranks to senior business operations analyst of the internal-facing Enterprise Technologies department.

Senior Business Operations Analyst

Served as internal Salesforce development liaison across 20 divisions and 5 offices of the firm to help build solutions by understanding changing business needs. Researched workflows and business processes from both end-users and leadership. Gathered requirements and prioritization for requests around new functionality.

Analyzed and monitored Salesforce licensing allocation, saving the company \$250k by optimizing the existing \$10M licensing agreement for a rapidly growing user base as the company expanded.

Collaborated with system architects and on- and off-shore developer resources to provide input and recommendation on design of functional solutions.

Created tailored Salesforce dashboards and reports, as well as running pivot tables and analyses in Excel.

Coordinator

Supported a team of 8 executives & 80+ staff with event planning, training, and project planning.

Served as project manager for a training course in a Crimson performance technology product that aimed to improve hospital performance for 150+ clinicians, physicians, and quality analysts.

Trained and on-boarded new staff.

Genetic Alliance, Programs Intern, Washington, DC, March – December 2013

Genetic Alliance is a DC-based non-profit dedicated to advocacy on behalf of communities affected by genetic conditions. I interned for their Baby's First Test division: a national clearinghouse of information related to newborn screening, a public health service.

Authored a series of 25 health blogs on a multitude of genetic conditions.

Compiled and edited a monograph of patient testimonials and an informational prenatal testing guide for expecting parents.

Authored a perspective on how whole-genome sequencing could be leveraged to better track the spread of infectious disease through a population.

Buchanan Elementary School, Teaching Aide, Lancaster, PA, January – May 2012

During the spring of my senior year of my undergraduate studies, I participated in the Science Teaching Internship as a course through my school Franklin & Marshall College. I was placed at Buchanan Elementary School and taught various topics weekly.

Taught a 3rd grade class (27 students) using my own inquiry-based lesson plans.

Taught a 5th grade class (24 students) using Full Option Science System (FOSS) curricula to teach geology and graphing.

Brain and Spinal Cord Institute, Research Intern, Paris, France, October – December 2011

During the fall of my senior year of my undergraduate studies, I participated in the Internships in Francophone Europe program through my school Franklin & Marshall College. I was placed

Performed transgenesis, creating transgenic lines of zebrafish by injecting embryos with plasmids of interest; screened transgenic lines using fluorescence expression patterns, and tested an algorithm to quantify their locomotive patterns.

Publications

Farrell, W. J.; Nair, D. T.; Miller, B. M.; & Zigler, A. M. 2016. The emergence and behavioral stability of social status in green anole lizard (*Anolis carolinensis*) dyads. *Animal Behavior and Cognition*, 3(3), 180–197. doi: 10.12966/abc.06.08.2016

Nair, D. "Whole-Genome Sequencing and Infectious Disease: A Novel Application of Sequencing Technology." Genetic Testing and Molecular Biomarkers. October 2013, 17(10): 719-720. doi:10.1089/gtmb.2013.1549.

Puffenberger, EG; Jinks, RN; Sougnez, C; Cibulskis K; Willert RA; et al. 2012. Genetic Mapping and Exome Sequencing Identify Variants Associated with Five Novel Diseases. PLoS ONE 7(1): e28936. doi:10.1371/journal.pone.0028936

Skills

Familiarity with local, state, and federal data sources and variety of data formats.

Software – R, PowerBI, Shiny, RapidMiner, Tableau, UCINET, Netdraw, SQL, Salesforce, XLMiner, SQLite, MS Access, SSMS, SPSS, PostGres

R Packages – Dplyr, Tidyverse, Shiny, Statnet, ModelR, Magrittr, Purrr, Stringr, Lubridate, Broom, Readr, Tibble, Data.table, XML, GGplot2, Rvest, Reshape2, Caret, Tree, iGraph, Nnet, Glmnet, Leaflet, SoilDB, Haven, TidyR, Sp, Sf

Models – Linear Regression, Classification, K-Nearest Neighbors, Regression Trees, Ridge/LASSO, XGBoost, Support Vector Machines, Neural Net, Random Forest, Principal Components Analysis (PCA)

Relevant Coursework

Business Insights through Analytics
Predictive Analytics
Database & Big Data
PowerBI: Dashboard in a Day

Organizational and Social Analytics
Business Process Management
Data Science
Text & Visual Analytics