

Bibor Szabo

Data Engineer | Data Scientist

Vienna, VA | 615.525.6826 | szabo.bibor@gmail.com | [LinkedIn](#)

OBJECTIVE

Data engineer with a background in statistical data analysis looking to leverage 4+ years of experience in a value driven environment.

SKILLS

- **Inference:** data processing, qualitative and quantitative data analysis, hypothesis testing.
- **Prediction:** data mining, supervised and unsupervised machine learning, predictive modeling, regression, classification, natural language processing/ NLP, data visualization, recommender systems.
- **Tools:** Python, Java, PHP, Streamsets, SQL, Plotly, Seaborn, Tableau, SPSS.

RELEVANT EXPERIENCE

Jr. Data Engineer | Jr. Engineer | Karsun Solutions LLC, Herndon, VA | 02/2021–current

- Earned the companywide shout out: “**from intern to key contributor within a year**”, as well as received the **Core Values Company Award** for my dedication to excellence, innovation, teamwork, integrity, and commitment.
- Automated end-to-end a data visualization dashboard that supports benchmarking efforts of 26 government agencies: from ETL through data processing to data visualization and display (Tech Stack: Java / Spring Boot, Python / Flask / Plotly, PHP / Drupal).
- Supported end-to-end integration and maintenance of a complex data application that automates data visualization to provide insight into Federal IT spendings for governmental use as well as for the general public.
- Assisted the cross-platform migration of multiple cloud based services from VPCaaS to a container-based MCaaS system.
- Created and maintained 80+ ETL pipelines in StreamSets to cleanse, pre-process, and migrate JSON data from REST endpoints to AWS S3 buckets on a set schedule.
- Indexed documents in ElasticSearch to enable the document and advanced search functionalities of a government website.
- Built 15+ Plotly visualizations to serve 6+ dynamically updating dashboards.

Hack the Bay | Chesapeake Bay Water Quality Hackathon | 08/2020–09/2020

- In an effort to aid the Chesapeake Bay Program, partnered with two data scientists to build and fine-tune a Random Forest classifier that predicted dangerous levels of phosphorus across the Chesapeake Watershed.

Data Science Immersive Program | General Assembly, Washington, D.C. | 03/2020–06/2020

- Completed 5 projects and 20 labs in a 500-hour full-time course, and presented my work to both technical and non-technical audiences.
- Using the Pushshift API, Natural Language Processing, and logistic regression, predicted the category of a social media post with 94.5% accuracy.
- Analyzed and visualized trends, used feature engineering, and fitted over 15,000 machine learning models to predict housing prices from a large number of features with the least possible error.
- As part of a team effort, collected data from 3 different websites, used NLP to conduct sentiment analysis on COVID 19 and social distancing-related twitter posts, and analyzed daily changes in new coronavirus cases on the state level. To provide an easily understandable visual display, we used Tableau to create an interactive map.
- **Capstone** – Using the Doc2Vec model of the Gensim library, created a search engine that recommends federal contract opportunities based on topic similarity.

Research Assessor | Peabody Research Institute, Nashville, TN | 03/2016–05/2017

- As part of a team, I supported two projects that parallelly focused on identifying potential areas of enhancement in the public education system to foster longer-term academic performance among under-served pre-Kindergarten-aged students.
- Conducted full-day classroom observations of the learning and social-emotional environment, teacher attitude, and quality of instruction.
- Assessed the cognitive development of students, using the Peabody Picture Vocabulary Test, Peg Tapping, and the Woodcock-Johnson III. psychoeducational battery.

Survey Research Methods Seminar | Vanderbilt University, Nashville, TN | 01/2009–05/2009

- Used SPSS to conduct a multivariate regression analysis and hypothesis testing on the 2008 Nashville Health Study to assess the impact of acculturation and discrimination experiences on the mental health of immigrants.

EDUCATION

General Assembly, Data Science Immersive, Washington, D.C. | 03/2020–06/2020

Middle Tennessee State University, M.A. in Sociology, Murfreesboro, TN | 12/2007

Master's Thesis - Acculturation Patterns of a Hungarian Sojourner Group Residing in Nashville, TN (qualitative analysis)

Eötvös Loránd University, J.D., Budapest, Hungary | 06/2002