VINAYAK SIVA KUMAR

1301 University City Blvd, Blacksburg, VA 24060

📞 540-998-0328 ✉ sivakumarvinayak@gmail.com 🔗 linkedin.com/in/vinayak-skumar/ 🔗 github.com/Vinayak-96

Education

Virginia Tech   **Aug. 2021 – May 2023 (exp.)**
Master's in Computer Science, GPA 3.34/4   *Blacksburg, VA, USA*
<u>Courses</u>: Data Structures and Algorithms, Data Analytics, Information Storage & Retrieval, Information Visualization, Advanced Machine Learning, Software Design & Quality, Big Data Engineering

Vellore Institute of Technology (VIT University)   **July 2015 – April 2019**
Bachelor's in Computer Science   *Vellore, Tamil Nadu, India*

Technical Skills

Programming Languages: Python, C++, C, R, Unix Shell Scripting
Databases and Big Data Tools: Databricks, AWS, MySQL, Oracle, Apache Spark, Hive, Airflow, T-SQL, Azure, GCP
ML & Data Libraries: Pandas, NumPy, Matplotlib, Scikit-learn, Tensorflow, MLlib, PyTorch, XGBoost
Reporting Tools: Tableau, Excel, Metabase, R Markdown
Web Technologies: HTML, CSS, JavaScript

Experience

Sleep Number Labs   **May 2022 – Aug. 2022**
Data Platform Engineer Intern   *San Jose, CA, USA*
- Built software components on Databricks Lakehouse to transfer and analyze biometric data, using Spark for data processing tasks and Databricks SQL dashboards for visualizations, which helped improve smart bed features.
- Refactored existing MLOps workflows with Databrick's MLFlow pipeline templates.

Tredence Analytics   **June 2019 – June 2020**
Software Engineer, Data Engineering Team   *Bengaluru, Karnataka, India*
- Created end-to-end data pipelines that collected customer and retail data from on-premise Hadoop clusters and migrated it to Google Cloud Platform (GCP) using Apache Airflow, which was then consumed as aggregated data in a reporting tool or fed into a propensity model for making customer behavior predictions.
- Designed Azure Data Factory (ADF) pipelines with Data Flow mappings and Databricks notebooks when a client needed to migrate their existing pipelines to a newer version of ADF.

Center for Study of Science, Technology and Policy (CSTEP)   **May 2018 – July 2018**
Intern, AI Team   *Bengaluru, Karnataka, India*
- Performed Data modeling and prediction of child malnutrition (with statistical inference and machine learning) present in rural districts of Karnataka, using census data and data collected from health care centers. The data was stored and processed as data warehouses and OLAP cubes using Pentaho's suite of BI tools.

Projects

Topic Modeling Analysis on Virginia Tech's Research | *PySpark, Spark NLP, MLlib, NLTK*   **Dec 2022**
- Developed a Latent Dirichlet Allocation (LDA) topic model that automatically identifies upcoming areas of research within the Computer Science department of Virginia Tech, using existing research papers available on VT's repository.
- The research papers were web scraped from the repository and stored on the HDFS of a Hadoop cluster, which was then fed into an NLP pipeline for text processing using spark-nlp, and finally sent into a topic model with Spark MLlib.

Virginia COVID-19 and Social Vulnerability Index Dashboard | *HTML, CSS, Javscript, D3*   **May 2022**
- Designed a website using Javascript and D3 which displayed interactive and custom graphs that captured information about COVID cases and the Social Vulnerability Index (SVI) identified in Virginia counties, to determine new social patterns and a possible relation between these two social factors. (https://infoviz-cs5764.web.app/)

Detecting Toxicity with Deep Learning (Undergrad Final Year Project) | *Python, NLTK, Tensorflow*   **April 2019**
- Created an online toxicity multi-label classifier which was able to parse words and classify harmful texts from social media into separate categories (such as racist comments, threats, insults) using Bidirectional Gated Recurrent Unit Recurrent Neural Networks with Stochastic Gradient Descent optimization, and was able to achieve an accuracy of 98%.
- Presented undergrad final year project in the 3rd International Conference on Intelligent Computing and Communication 2019, Springer and it was later published as a book chapter. https://doi.org/10.1007/978-981-15-1084-7_57

Certifications

- Databricks Academy Accreditation - SQL Analyst Associate (June 2022)

- Indian Institute of Technology, Tirupati Summer School on Machine Learning 2021