Avinash Ramesh

San Jose, CA | (408) 338-5862 | avinash.ramesh@sjsu.edu | www.linkedin.com/in/avinash94/ | Github: github.com/rameshavinash94/

SUMMARY

A Software Engineer with over 5 years of hands-on experience in the backend/data engineering space, as well as a strong desire to work on end-to-end AI/ML products that can scale and aid in ML model productionization.

EDUCATION

San José State University *San Jose, CA, USA*
Master of Science in Computer Software Engineering (Specializing in Data science) ***August 2021 - April 2023***

Anna University *Chennai, TN, INDIA*
Bachelor of Technology in Information Technology ***June 2012 - April 2016***

EXPERIENCE

Carl Zeiss Meditec, *Dublin, CA* ***May 2022 - December 2022***
Software Engineer Internship - Building data products to support AI/ML analysis

- Spearheaded the entire product development lifecycle for **three** key functionalities of CRUD-based DICOM application: uploading series of scans, annotating studies, comprehensive metadata search and exporting dicom metadata & scans.
- Collaborated with product managers and downstream users to plan, design and develop an internal data product within 3 months of internship to fasten clinical research and analytics for AI teams.
- Follow object oriented design patterns and dependency ingestion techniques to develop loosely coupled applications.
- Built event-driven data pipelines, scalable APIs handling nearly **100 concurrent users**, API pagination, deployed database models, and developed a cache system for storing/retrieving key value pairs and preventing database connection exhaustion.

ITIDATA, *Chennai, India*
Associate Data Engineer - Big Data ***July 2019 - July 2021***

- Planned infrastructure to process internal & external vendor data feeds for global use within Citigroup and apply business rules & data selection hierarchy on individual data attribute levels to create a golden data source.
- Migrated Ab Initio graphs/plans to Spark distributed framework (POC) nearly **40%** faster thereby reducing licensing costs spent on ETL tools.
- Identified, reviewed, and implemented internal process improvements: automating manual processes, optimizing data delivery by **50%**, re-designing infrastructure for greater scalability, etc.
- Created multiple Hive tables with partitioning and bucketing for efficient data access; optimized existing SQL queries by **30%**; and wrote Sqoop jobs to import, export, and update data between HDFS, Hive, and relational databases.

Analyst / Data Engineer ***August 2016 - June 2019***

- Implemented ETL pipelines to extract more than **2 million** records on a weekly basis from various vendors(Reuters, Bloomberg, etc.), wrangle ingested data, apply business transformation rules, and load into desired formats.
- Created a framework to recognize data anomalies and produce daily automated reports to reduce turnaround time to nearly **half** for addressing data problems.
- Designed database schemas and applied Slowly Changing Dimensions(SCD) in Data Warehouse to track historic changes.
- Remodeled functional logic of python similarity script using Numba jit, numpy arrays and map operations resulting in decrease in time complexity by nearly **3/4th**.

SKILLS

Programming languages: Python, Unix Shell Scripting, SQL, Java, Javascript **ETL tools**: Ab Initio, OpenRefine
Big Data Frameworks/Tools: Spark, Hive, Hadoop, HDFS, YARN, Sqoop **Scheduler**: Autosys, Crontab
Databases: Oracle 11g & 12c, SQLite, HBase, Postgresql **Data Formats**: CSV, JSON, Parquet, Avro, DICOM
Web Frameworks/API: HTML, CSS,React, Flask, Fastapi **Visualization**: Tableau, Microsoft Excel, Streamlit
CI/CD tools: IBM UrbanCode Deploy, AWS code pipeline, Azure Devops **Version Control**: Bitbucket, Github
Container Technology & Management, Configuration: Docker, Kubernetes, Ansible
Cloud: **Azure** - Azure Blob, App Service, Azure Functions, Azure Data Factory, Azure Batch, Azure Devops, App Logic
AWS - EC2, S3, ELB, LAMBDA. **GCP** - Cloud Run, Cloud Build, GKS, Vertex AI
ML Libraries: Numpy, Pandas, Matplotlib, Pyspark, Spacy, Scikit-learn, Keras, Tensorflow, Pytorch
DS Concepts: Classification, Regression, Clustering, Dimensional Reduction, NLP, Ensemble, DL

ACADEMIC PROJECTS

Status : *Completed* ***September 2021 - December 2021***

- **Wikipedia Based Question & Answering (QA) Application**:
  Repo: *https://github.com/rameshavinash94/Wikipedia-QA-System*
  Techniques**:** *Information Retrieval, Cosine Similarity, Word/Sentence Embedding, QA systems, BERT, NLP*
  Libraries**:** *Python - Spacy, Transformers, Wikipedia-API, Streamlit, Pandas, Numpy*

- **Web App for Detection & Classification Of ECG Images**:
  Repo: *https://github.com/rameshavinash94/Cardiovascular-Detection-using-ECG-images*
  Techniques**:** *Image processing, Dimensionality reduction, Ensemble*
  Libraries**:** *Python - Scikit-Learn, Scikit-Image, Matplotlib, Pandas, Numpy, Joblib, Streamlit*

- **Patient Management System (Book/Cancel Appointments with Doctors):**
  Repo: *https://github.com/rameshavinash94/CMPE272_PMS*
  Libraries**:** *Python - Flask, Flask-oidc, okta, Dialogflow, sqlite Web - HTML, CSS, JS, Bootstrap*

Status : *Completed* ***January 2022 - May 2022***

- **Flight Booking Application:**
  Repo *: https://github.com/gopinathsjsu/individual-project-rameshavinash94*
  Techniques*: OOP principles and design patterns: state pattern, chain of responsibility pattern*
  Language and Framework: *Java, maven*

- **Alternus Vera - User Intent - Intention based ML/ Deep learning approach for fake news detection**
  Repo: *https://github.com/rameshavinash94/Alternus-Vera*
  Techniques: *Data Amalgamation, Data wrangling, Deceptive cues, NLP processing, Micro Factors, semantic cues, Deep Learning*
  Libraries: *pandas, seaborn, matplotlib, nltk, spacy, transformers, numba, spacy_universal_sentence_encoder, tensorflow*

- **Real Estate Analysis**
  Repo: *https://github.com/rameshavinash94/Real-Estate-Analysis*
  Techniques*: EDA, Data wrangling, finding micro factors contributing to real state prediction, Data amalgamation, Classification, CLustering, Regression, Dashboard*
  Libraries*: Python, pandas, numpy, matplotlib, BeautifulSoup, scikit learn*

- **Deep Learning Framework for Facemask Detection using CNN, Facemask-Removal using GANs and Gender Classification**
  Repo*:https://github.com/rameshavinash94/Deep-Learning-Framework-for-Facemask-Detection-CNN-Facemask-Removal-GAN-s-and-Gender-Classification*
  Techniques*: GAN, Facemask Identification, MLOPS, Deep Learning for Gender Classification*
  Libraries*: cvlib ,pandas, numpy , streamlit, joblib, tensorflow, opencv, matplotlib*

- **Retail Store Analysis**
  Repo*: https://github.com/rameshavinash94/Retail-Store-Analysis*
  Techniques*: EDA, Data wrangling, finding factors, Data amalgamation, Classification, Regression, Clustering, Dashboard for visualization*
  Libraries*: Python, pandas, numpy, matplotlib, BeautifulSoup, scikit learn, pycaret*

- **Company Analysis - Knowledge Graph**
  Repo*: https://github.com/rameshavinash94/Knowledge-Graph*
  Techniques**:** *Web Scraping, NLP stages of operations****,*** *Data Extraction and Amalgamation for Knowledge Graph building*
  Libraries*: Python, pandas, numpy, matplotlib, plotly, BeautifulSoup, py2neo, spacy, tensorflow, neo4j*

- **Poem Generation**
  Repo: *https://github.com/rameshavinash94/POEM-NLP/*
  Techniques: *Web Scraping, NLP operations (POS, ComputeSimilarity, preprocessing), visualization*
  Libraries: *pandas, numpy, tensorflow, numba, BeautifulSoup*