Shoban K.

Email: shobankumarrv@gmail.com

UIUC Research Park
Mobile: +1-765-772-5154

Summary

Diligent software engineer with over 5 years of experience with expertise in developing back-end services for distributed systems, machine learning applications with high scalability and low latency. Able to effectively self-manage during independent projects as well as collaborate as a part of a productive team.

Education

Purdue University — West Lafayette
Master of Science in Computer Science — Dec 2020
Courses: Statistical Machine Learning, Deep Learning, Data Mining, Cloud Computing, Analysis of Algorithms

Anna University — Chennai
Bachelor of Technology in Information Technology — May 2015
Courses: Data Structures and Algorithms, Cloud Computing, Web Development, Operating Systems

Skills Summary

Languages	Python, JAVA, JavaScript, C++, SQL, Bash
Frameworks	Scikit, NLTK, SpaCy, TensorFlow, Keras, Django, Flask, NodeJS
Tools	Kubernetes, Docker, GIT, Maven
Platforms	SAP HANA, AWS, GCP
Core Strengths	Leadership, Event Management, Debugging, Data Analytics, Time Management

Experience

Yahoo! — Champaign
Software Development Engineer — March 2021 - Present

- **Financial Data Warehouse (Hadoop, Pig, Oozie, Java, Python, Looker)** Part of Yahoo's data lake team which consolidates the revenue and payment data of entire Yahoo ecosystem and manage as a central source of truth for revenue forecasting and growth across multiple platforms.
- Worked on an adaptive re-agggregation framework which helps in faster querying and reducing the cost by more than 20%
- Part of the oncall cycle responsible for managing and monitoring issues which impact customers with highest quality
- Providing analytical insights to marketing and sales team with improving the targeted revenue for customers by tens of thousands of dollars

Yahoo! — Champaign
Software Engineering Intern — June 2020 - Aug 2020

- **Report Summarization Framework (JAVA, Python, Ember.js, NLP & Machine Learning - Augmented Analytics)** Designed and developed a new framework to provide automatic insights on the report generation and summary
- Created real - time prediction of the report data on anomaly detection and trend forecasting using statistical approach
- Reduced the manual effort of the Business Analysts and Product Managers by more than 30% during report investigation and analysis
- Generated new automated insights and reports using templates and thresholds which help to uncover potential revenue and data quality impacts

Purdue University - Agricultural & Biological Engineering — West Lafayette
Graduate Research Assistant — Mar 2019 - Dec 2020

- **Gene Regulatory Networks (Python, Scikit, Statistical Analysis, Machine Learning and Deep Learning)** Worked on Machine Learning Algorithms using Classification Tree Analysis approaches to improve the mapping between the plant genes by 50%
- Researched statistical and heuristics approaches for large scale high dimensional data to enable parallelism in the data pipeline by reducing the runtime of data processing by more than 40%
- Developed a new algorithm to enable and identify mapping between genes of different plant species and help us to understand the relationship between multiple them which helps the biologists to reduce time and cost for their experiments

- Created jobs and schedulers on clustered environment to aid multiple students in their structured analysis of their results using latest techniques

Purdue University - Department of Computer Science West Lafayette
Graduate Teaching Assistant May 2019 - Jul 2019

- **CS 240 (Programming in C)** Graded exams for about 60 students and consistently delivered one-to-one assistance during lab sessions
- Provided support and clarification of doubts for students during office hours

SAP Labs Bengaluru
Software Developer Nov 2015 - Jan 2019

- **Document Processing and Analysis(Javascript, Python, NLP)** Worked on the Document Processing pipeline which processes the unstructured data from the external systems into meaningful insights and put into the Clinical Data Warehouse
- Developed Document Processing Workbench which simplifies the ingestion of the unstructured data from the external systems, easier assignment of the pipelines and helps in better visualization of the meaningful insights extracted
- Developed Document Processing Workbench which simplifies the ingestion of the unstructured data from the external systems, and helps in better visualization of the meaningful insights and reduced the manual efforts by more than 40%
- Maintained production code by writing unit tests, scenario tests and fixing issues in the Acceptance testing phase
- Helped and mentored 3 junior software developers to be upto speed with the product and technical knowledge

VMWare Bengaluru
Associate Software Developer Jul 2015 - Nov 2015

- **VMWare WorkSpaceOne(Javascript, C#)** Worked on the console team that helps the Admin to maintain thousands of devices efficiently
- Assisted in identifying new issues and fixing the production issues for various platforms

Projects

Fake News Detection (CS 573 - Data Mining) Classified news headlines from 2015 U.S. Presidential Elections into real or fake news by designing an LSTM and evaluating it against common classification and regression algorithms from scikit-learn. Developed weight based technique to find the factors affecting fake news(model achieved an accuracy of 96%)

Adversarial detection against Neural Networks (CS 590 - Program Analysis for Deep Learning) A deep learning approach to generate adversarial samples from the MNIST dataset and check the robustness of multiple Convolution Neural Network models and proved that our novel approach can detect the adversarial samples with better accuracy

Image Sharing and Classification(Research Project on Machine Learning and Cloud Computing) A micro service-based framework which allows the users to upload millions of images to the database and effectively classify them using ResNet. The serveless architecture was designed with low latency, fault tolerance and secure acccess gateways.

Shape Detection (ECE 695 - Deep Learning) Applied noise reduction techniques including Bilateral Filtering and Variational Auto Encoding on noisy image data with complex shapes and structures.

Secured Indexing in Encrypted Cloud Data Proposed an encryption based model for indexing any documents in the cloud database. This enabled faster search without the cloud provider knowing any information related to the documents using a hybrid combination of hashing and encryption. Also, prefixed the indexing with trapdoor function which makes nearly impossible for adversary to gain any details

Chat Bots for Auto-responses A hybrid approach using generative and retrieval based approach to generate automated response to the questions using Chat Bots on sports and science domains. The challenging part was to generating response to the text which was handled efficiently using combination of Bag of Words , TF-IDF Vectorization / Cosine Transformations