Manav Bagai

mbagai@asu.edu| (480)-570-8392 | https://www.linkedin.com/in/manav-bagai/ | https://manavbagai.github.io/

EDUCATION

Arizona State University, Tempe, USA **August 2018- May 2020 (Expected)**
Masters in Computer Science GPA: 3.94/4
Aligarh Muslim University, Aligarh, India **August 2012- June 2016**
Bachelor of Technology in Computer Engineering GPA: 8.1/10

SKILLS

Programming Languages: Python, Java, Scala, Groovy
Cloud Platforms and Dev-Ops: Azure, AWS, GCP, Jenkins, Docker, SaltStack, Linux, Ansible
Data Engineering: Airflow, MySQL, Neo4J, Druid, Superset, Hadoop, Spark, Nutch, Kafka, Redis
Web Development: Play Framework, Flask, Django
Machine Learning and NLP: Scikit-Learn, Stanford CoreNLP, WordNet, Tensorly, Tensorflow

EXPERIENCE

Data Engineer Intern, Starbucks, Phoenix, USA **May 2019- Present**

- Developed a **Real-Time** Social Media Analytics Platform for Starbucks using **Python**, **Druid**, **Kafka**, **Superset**, and **Docker**.
- Worked with the Security team to find vulnerabilities in **Azure** based applications by querying various agents like **Nessus Agent**, **OMS Agent**, and **Splunk** Forwarders.
- Currently working with the CI/CD team to deploy automated solutions using **Azure**, **Jenkins**, **Ansible**, and **Kubernetes**.
- Working on a POC to automate the vulnerability patching process of **Azure** based resources.

Big Data Engineer, Exadatum Software Services, Pune, India **November 2016- February 2018**
Projects:
Ingestion, Analysis, and Visualization of Cardiovascular Patient Data **June 2017- February 2018**

- Developed an ETL pipeline in **Apache Spark** which is used for ingesting the data in algorithm that computes the Risk Score.
- Orchestrated the data flow using **Apache Airflow** and integrated it with a web application in **Play Framework**.
- Developed and Optimized a **Docker** Image with **Hadoop stack** installed to make the above system ready to run and deployed.
- Worked on deploying and running the above application on **Stanford VM** and client **AWS** environment.
- Responsible for creating and maintaining client environment using **AWS, Docker Images, Jenkins, Databases** and **SaltStack**.

Recommendation System and Chat-bot **November 2016- May 2017**

- Created the knowledge base by crawling customer support data from Kohl's and Macy's website using **Apache Nutch**.
- Cleaned the data and generate required csv and json using **Python** and **Pandas**.
- Ingested the data in **Neo4J** and written parser to generate hierarchal json to be ingested in **API.AI** for chat-bot.
- Written **REST** web application in **Play Framework** to query **Elastic Search** and send data to the recommendation system UI.

ACADEMIC PROJECTS

Save the Underprivileged Children **March 2019- May 2019**
Technologies: Spring, React JS, Docker, AWS- S3, Rekognition API, GCP- Google App Engine, Cloud SQL

- The main aim of the project is to help underprivileged children by reducing the gap between NGOs & underprivileged children.
- Any end-user can upload the image of the unprivileged child which is shared with NGOs and NGOs can assign volunteers.
- A face image-based database is created using the Face Recognition feature of Amazon Rekognition API and matches the face with already registered children to check if the child is a new entry.

Video Surveillance Service- Multi-Tier Cloud Based Application **January 2019- March 2019**
Technologies: Play Framework, AWS- EC2, Snapshots, S3, and SQS

- Web Tier is responsible for to get the request from Client and add it to Amazon SQS.
- App Tier is responsible to run the Deep Learning Model to detect objects and add the results to S3.
- Web Tier also acts as a load balancer i.e. based on the number of requests in SQS, App tier instances are launched.

Activity Recognition System **August 2018- December 2018**
Technologies: Python, Numpy, Pandas, Scikit-Learn

- Developed a system to recognize daily activities such as eating and sleeping by applying feature extraction, feature selection (PCA), and classification (SVM, Decision Tree, & ANN) on time series data captured using Myo sensor.

ACHIEVEMENTS

- **Stood 1st at University of Arizona- HACK ARIZONA hackathon (36 hours)** **January 2019**
  Part of a 4-member team- built IRIS (Ideal Recruiting, Intelligent Solution) which is an AI based framework for recruiting.
- **Certified Neo4J Developer** by Neo Technologies- Founder of Neo4J Graph Database.