Sen Han

Looking for a Ph.D. position in computer science

Email: senyan@outlook.com | Mobile: (+1)320-339-7110

EDUCATION

St. Cloud State University **St. Cloud, MN**

M.S. in Computer Science Jan. 2011 - Jun. 2013

Core Courses: Evolutionary Computation, Advanced Operating System, Computer Architecture, Distributed Systems

Publication: Batting Order Optimization by GA published to GECCO ACM

B.S. in Computer Science Sept. 2007 - Dec. 2011

Core Courses: Evolutionary Computation, Probability, Discrete Math 1 and 2, Calculus 1 and 2

WORK EXPERIENCE

Bloomberg Industry Group **Washington, DC**

Technical Manager Aug. 2019 - Present

- **Team Management and Leadership:** Supervised a team of 15 engineers for both data engineering and web app development on daily basis including requirement analysis, coding, refactor, quality assurance; Founded the group's technical interview panel for technical excellence; Co-Founded cross-department cloud committee for AWS cloud migration governance and technical excellence
- **Machine Learning:** Pivoted the business's human entry based data enhancements into a machine learning approach which leads to multi-million dollars strategic business investment; Initiated several projects including Pwin (Probability of Win) Forecast for the government contract, entity extraction from government documents such as security clearance requirements and place of performance as well as the recommendation system for government opportunities
- **Data Engineering:** Initiated a reusable components based approach to revamp the company's 11-year-old data pipeline with Spark-based data lake for hundreds of millions of government transactions and terabytes of government documents as the principal architect and leader for the group's AWS migration for both application and data pipeline

Senior System Architect Jul. 2017 - Aug. 2019

- **R & D:** Led initiatives including Kubernetes orchestration on Bloomberg OpenStack private cloud; Designed and Implemented core email campaign tool, topic-based news alerts systems; Transformed the team's system into a container-based modern system which increases the daily deploy to multiple deploys per day; Achieved greater stability and reliability with unified logging monitoring and alarming
- **Product Deployment:** Kubernetes cluster; Email campaign tool; topic-based news alerts; Jenkins CICD pipeline

Bloomberg LP **Washington, DC**

Senior Software Engineer Jul. 2013 - Jul. 2017

- **R & D:** Worked with domain experts and product managers to identify customer needs in their workflow and provide the innovative technical solutions by using a Bloomberg style data-centric approach such as web application development, data acquisition and data indexing with Elasticsearch
- **Product Deployment:** Government Contract Search; Opportunity Search and Pipeline Manager; Keyword Suggestion

SELECTED PROJECTS

Recommendation System of Government Opportunities **Bloomberg, Washington, DC**

Technical Leader Jan. 2021 - Present

- **Problem Investigation:** Investigated the in-site opportunities recommendation system, such as user dwelling time and click through rate; Gathered user feedback through anonymous survey which shows that the opportunity recommendation does not meet the expectation
- **Model Design and Development:** Designed and developed a user interests model based on the users' long-run interests, specifically leveraging TF/IDF model computed on the users' public award history; Categorized the opportunities to several topics; Built a new recommendation system based on user interests
- **A/B Test: Deployed** A/B tests and designed several metrics to measure the gain of treatment over baseline, which shows that the new recommendation system has a higher user engagement and retention rate
- **Deep Learning-based Model**: Designed and trained a BERT based siamese model as a POC, which vectorizes the user and opportunities to 100D vectors and recalls opportunities based on cosine similarity of user embedding and opportunity embedding

Document Information Extraction **Bloomberg, Washington, DC**
Technical Leader Mar. 2020 - Present

- **Entity Extraction System**: Built a name entity extraction system based on Spark NLP to extract entities from documents, such as place, key personnel and clearance requirements
- **Data Preparation**: Designed and developed a pipeline to facilitate human dataset annotation to gather more high-quality training data
- **Model Fine-tune**: Implemented an exhaustive search method and model traceability framework to try out different combinations of data preprocessing methods and model hyper-parameters
- **Project**: Held brain-storming session in the team to bring up new ideas; Worked with business stakeholders to pivot from a labor-intensive approach to a ML approach; Piloted the initial POC implementation which transformed into multi-millions strategic investment in next fiscal year

Probability of Win **Bloomberg, Washington, DC**
Technical Leader Mar. 2019 - Feb. 2020

- **Data collection and normalization**: Carefully designed a spark based data pre-processing framework on AWS to handle a large amount of unstructured data, such as public government transactions of a given vendor, and normalized the data into a standard format
- **Model Development**: Adopted fussy clustering and hierarchical clustering algorithm to cluster the normalized data, and then computed the probabilities of winning the bid based on the project clustering and the vendor's past award history

Data Lake Project **Bloomberg, Washington, DC**
Chief Architect Jan. 2019 - Present

- **Data lake development**: Transformed a legacy Oracle PL/SQL based data pipeline into a modern Spark-based data lake on AWS

Keyword Suggestion for In-site Search Engine **Hackathon Project, Bloomberg**
Project Owner of Winning Hackathon Project Jun. 2014 - Dec. 2014

- **Data Preparation**: Built easy-to-use data pipeline to extract training and evaluation data from PDF and Word documents by leveraging techniques like OCR model; further developed dataset cleansing pipeline to filter out invalid data, like de-duplication and typo fix
- **Exact Match-based Method Development**: Leveraged the full-text search function provided by Elasticsearch, which uses BM25/TF-IDF, indexed the extracted phrases from the large dataset, and then built an auto-complete suggestion system based on Elasticsearch
- **Deep Model-based Method Development**: Studied word and sentence embedding-based methods, such as Word2Vec and Doc2Vec, finetuned a BERT-base model by taking the idea of sentence and phrase vectorization to vectorize the extracted phrases, and then combined cosine-similarity with hand-crafted features to do keyword suggestion with a multi-layer perception network

SELECTED REWARDS

Engineering Award of Excellence, Bloomberg Industry Group **Washington, DC**
Quarter winner 1 out of 200 engineers 04/2021

Management Award of Excellence, Bloomberg Industry Group **Washington, DC**
Quarter winner 1 out of 5000 employees 05/2019

Hackathon Winner - Market Impact, Bloomberg LP **Washington, DC**
Winner 1 out of 35 cross-functional teams 07/2013

CERTIFICATES

AWS Certified Solution Architect - Professional 09/2019 - Present
AWS Certified Security Specialty – Professional 10/2019 - Present
Linux Certified System Administrator - Linux Foundation 10/2018 - Present
Openstack Certified Administrator - Linux Foundation 12/2018 - Present
Kubernetes Certified Administrator - Linux Foundation 11/2017 - Present

SKILLS

Programming: C/C++, Java, Python, Ruby, Typescript/Javascript
Software & Tools: Kubernetes, Openstack, Vim, Maple, Matlab