Aniruddha Kalburgi

ORACLE Certified Professional Java SE 7 Programmer

Mobile: +353 (0) 899647136 Ireland
LinkedIn: https://in.linkedin.com/in/anikalburgi **Email:** andy-k19@outlook.com

Experienced Machine Learning Engineer & Backend Software Developer
- 9 months in ML, NLP – text classification, custom entity extraction, documents clustering and classification, rule based Data Mining
- 3 years of experience in Backend Software Development- Java, Microservices, Spring Boot, MongoDB, DB2, JMS, JUnit, Cucumber
- Masters in Artificial Intelligence from CIT, Cork - 2.1 Honours Degree, 2019

Technical Skills:

Artificial Intelligence, Machine Learning, Deep Learning, NLP, Recommender Systems, Big Data Processing

AI/ML/NLP: Keras, Tensorflow, Scikit-Learn, NLTK, spaCy, Gensim, PySpark, NumPy, Pandas, FastAI
Programming: Python 3, Java 8, Shell Scripts
Frameworks: Spring Boot, Spring Data, Hibernate, JUnit, Cucumber, JMS
Tools: Git, Docker, Jenkins, Maven, Gradle, JWT, Jetbrains IDE's, Profiling tools
Database: MongoDB, DB2, MySQL
Cloud: OpenShift, Databricks
Domain Knowledge: Forex Settlement & Payment Gateways (SWIFT, SIC, FED), US Retirements Investment Plan Customization (401k, 408k, 457b), TDD, BDD, Agile Development, Hedge Funds & Mutual Funds

Work Experience: (Total 3.9 years of experience)

Jul 2019 to Present **Contractor- NLP/ML Developer** **Clearstream, Cork, Ireland**

Using Data Mining, Machine Learning and NLP techniques to extract data from Unstructured Hedge Fund PDF documents.
Key Technologies: Python 3, Microservices, NumPy, Pandas; ML/NLP - Scikit Learn, Tensorflow, Keras, FastAI, SpaCy

- Fund Name Extraction using ML, NLP – TF-IDF, Decision Trees, SVM, LSTM based models.
  - Proved high importance of preserving symbols & NER, POS tags to identify fund names.
- *Self-Initiative* - Clustering visually similar documents using textual and engineered visual features – Agglomerative Clustering, spaCy, scikit-learn.

    - Demonstrated use case in identifying unique PDF templates per fund administrator, and possible production application to identify new templates.
- Documents Classification for 30+ document types – TF-IDF, SVM, Word2Vec, LSTM
- Labelling and Training of custom NER model Fund Name Entities Recognition using Prodigy from spaCy
- Fund Name Entities Extraction and Matching algorithm – Custom NER model used to improve entity extraction.
- Data Mining from Unstructured Contract Note PDF documents - using Python, ML
    - Fund managers (client) send PDF trade contract notes with no standard structure & use multiple templates; On OCR extracted data, we use regex rules to map certain values to fields, while language statements aren't straight forward to map, especially when they appear without label, so we offered improved solutions using Machine Learning and NLP to understand and map this additional information to correct fields.
    - Also doing PoC on automating complete information extraction using NLP and ML.

Dec 2017 to Aug 2018 Analyst – Software Developer TIAA GBS, Pune, India

- *Key Technologies:* Java 8, MongoDB, Spring Boot, Microservices, Mule ESB, JUnit, Cucumber-Java, Jenkins, Openshift, Docker, JWT.
- Newly created and enhanced the existing Java Microservices based applications using Spring Boot, Spring Data, MongoDB and more.
- Built BDD and Unit test cases using Cucumber, JUnit and Mockito.
- Designed some new, and updated existing Mule ESB workflows, configuration scripts for deployment on Jenkins Pipelines and Openshift cloud environments.
- Newly learned and implemented a JWT token based authorization mechanism.
- Outperformed in several sprints by fast delivering on the tasks while ensuring the best coding standards.

Sep 2015 to Nov 2017 Software Engineer Xoriant Solutions Pvt Ltd, Pune, India

- *Key Technologies:* Java 7, DB2, Spring, Hibernate, IBM Websphere MQ, JUnit.
- Worked on Financial Payments Message Formats like SWIFT SFin Messages (through SWIFTNet), FED (through Fedwire) & CHF - Swiss Franc (on SIC IGT Plus n/w).
- Core Payments Components - Owner & SPOC for sub-components that managed currency open and close, contingency and reconciliation events for over 18 Central Banks worldwide.
- Worked on enhancements and bug fixes for PayIn, PayOut, and caching algorithms.
- Revamped the system to adhere to the latest standards of the financial message exchange networks like SWIFTNet, SIC IGT and Fedwire.
- Helped DevOps and other teams to fix failures in Component Integration Tests and User Acceptance Tests.
- My team was showered with numerous appreciations by our client for completing all tasks efficiently and within deadlines, as well as continually upgrading our skills and implementing them in the project. For this, we won the 'Star Performer Team' award on multiple occasions.

Education:

Sep 2018 – May 2019	**MSc Artificial Intelligence – 2.1 Honours Degree** **Cork Institute of Technology, Cork, Ireland**	
Core Modules:	Practical Machine Learning Recommender Systems Fraud & Anomaly Detection Big Data Processing Research Practice & Ethics	Deep Learning Metaheuristic Optimization Decision Analytics Knowledge Representation
Sep 2012 – May 2015	**B.E. Computer Engineering** **Dr. D. Y. Patil Institute of Technology,** **Pune, Maharashtra, India**	
Sep 2010 – May 2012	**Diploma in Computer Engineering** **Dr. Bapuji Salunkhe Institute of Engg. & Technology,** **Kolhapur, Maharashtra, India**	

Projects Highlights:

Masters Research Project:
Recommending Diverse and Serendipitous Movies using Tag-genome
Tools & Techniques Used: Python 3, Scikit-Learn, Numpy, Pandas, SciPy, Unsupervised Learning, Stemming, Lemmatization, NLP, K-Means, DBSCAN & Hierarchical Agglomerative Clustering, Multi-Processing and Threading, PyCharm, Jupyter Notebook. Metrics for measuring Diversity and Serendipity of Recommendation List.

- **The objective** was to broaden the user's preferences by recommending then the Serendipitous movies. A serendipitous movie has high relevance to the user's taste, offers diversity in taste, a user wouldn't have discovered it on their own and user didn't expect to have that movie recommended to them, yet, the movie broadens user's taste.
- **Dataset:** Movielens dataset offers tag-genome information; Generated from the user ratings and movie tags from the Movielens dataset and public IMDB reviews. It describes all movies using the fixed 1128 tags-genomes. Values are decimal relevance scores between 0 and 1. e.g. fiction-0.1, romance-0.05
- **How:** Infer every user's taste based on tag-genomes for the movies they've watched, form clusters of tag-genome term vectors of movies, such that each cluster exhibits logical preference of the user. Optimal cluster size is chosen based on the highest Silhouette Score. Choose some top-ranked movies from each dense cluster - user's most preferred category while the remaining majority recommendations are chosen from the sparse clusters; using the custom novelty based reranking algorithm; which ranks movies in such a fashion that primarily focuses on Serendipity.
- **Results:**
  - Achieved 10% higher Serendipity for 435 real users from the known state-of-the-art Serendipity-sac2018 dataset.

- Proposed Novelty Re-Ranking Algorithm which is configurable through four parameters to manage Diversity-Serendipity trade-off.
- Also concluded that higher Diversity affects the Serendipity scores, where balancing similarity to the user-profile boosts them, linking tag-genomes using Lemmatization produces better quality clusters.
- Thesis, reports and source code links available here: https://www.researchgate.net/profile/Aniruddha_Kalburgi2

Other Mini-Projects:

- Text and Images classification using Deep Learning Neural Network Architectures.
- AI-based Battle-Tank agent game to fight off enemies and interact with the game world.
  - Multiplayer AI-based battle-tank agent teams, each tank positioned randomly, fight against opponent team tanks, by making their attack strategies, automatically finding the path to the enemy, while continuing to discover bonus ammo and smartly overcoming the hurdles placed at random places.
  - Knowledge discovered by the agent is made available to all agents from the same team. Each agent also tries to save themselves when health is low and attack only when necessary.
- Predicting IMDB score for movies from the dataset 'IMDB 5000 Movies'.
- Predicting Benign or Malignant class for Cancer dataset.
- Designed Genetic Algorithms to find optimal paths for the Traveling Salesman Problem.
- Implemented Metaheuristics algorithms to find the best solutions - Iterative Local Search with an enhancement of Novelty and Novelty+.
- Used PySpark to answer analytics related questions for datasets of Coca-Cola Bicycles and BusEireann bus schedules in Cork city.
- Used GeCode library to formulate CSP/CSOP (Constraint Satisfaction & Optimization Problems) to solve problems like a solution to the famous 'Einstein's Puzzle' using CSP, 'Building Projects Optimal Investment' using CSOP.

Achievements:

- *Hobbies:* Swimming, Table Tennis
- Oracle Certified Professional, Java SE 7 Programmer – Dec 2016.
- Winner in annual organization level programming challenge – TechX – 2016.
- Achieved 1st rank in district level Java Programming competition in India, 2012.
- Cleared level-2 ***musical instrument (Tabla) exam*** – 2005.