Sukhada Ghewari

sukhada.ghewari@gmail.com | linkedIn/sukhada-ghewari | github/sukhada14 | Portfolio

EDUCATION

University of California San Diego GPA: 3.97/4 | Sept 2021 - March 2023
MS in Computer Science (AI specialization)
Coursework: Algorithms, Principles in Computer Architecture, Recommender system and web mining, Data systems for ML, Data mining and analytics, Machine learning for physical applications, ML: learning algorithms

Walchand College of Engineering, Sangli GPA: 9.51/10 (Gold Medalist) | July 2015 - May 2019
B.Tech in Computer Science and Engineering
Coursework: Data Structures and Algorithms, Operating Systems, Distributed and Cloud Computing, Digital Image Processing, Internet and Web Programming, Computer Networks

SKILLS

• Python • Java • SQL • PySpark • PyTorch • Flask • Git • Data science toolkit- sklearn, pandas, matplotlib • Jupyter Notebook •NLP

WORK EXPERIENCE

ORACLE INC. | Software Development Intern Redwood City, CA | June 2022 – Sept 2022
- Developed an auto-healing system that recommends solutions to the errors in the CI/CD pipeline using Natural Language Processing techniques (TF-IDF vectorizer, stemming, lemmatizing) and Gradient boosting algorithm.
- Built an end-to-end Machine Learning pipeline by integrating the model in the CI/CD pipeline using Java, Spring Boot, Maven, ReST APIs, and OJET, which will be pushed to production.

IDEKER LAB, UC SAN DIEGO | Graduate Student Researcher La Jolla, CA | Oct 2021 - June 2022
- Built a novel Synapse Protein Ontology based on synaptic proteins interaction and isoform data.
- Improved the performance using Random Forest and community detection algorithms such as CliXo, HiDef to create the ontology and neural network architecture for psychiatric disease prediction. (FDR <= 0.05)

TCS RESEARCH | System Engineer Pune, India | July 2019 – May 2021
- Deployed a Python tool to detect bias in ML models using confusion matrix and fairness metrics.
- Authored a research paper and filed a patent through TCS in India domain for 'Domain Name Generation Algorithm' for defensive registration to curb phishing and other malicious attacks using CFG and unsupervised learning.
- Collaborated with research-focused team members and honed technical writing and project management skills.

ALGOANALYTICS PVT. LTD. | Data Science Intern Pune, India | Jan 2019 – June 2019
- Performed a comparative study and built semi-automated ICD-9 (International Classification of Diseases) Coding system using deep learning (F1 score: 0.51).
- Created an automating Question-Answer System using ensemble word embeddings and cosine similarity.

PUBLICATIONS

Kumar N, **Ghewari S**, Tupsamudre H, Shukla M, Lodha S. When Diversity Meets Hostility: A Study of Domain Squatting Abuse in Online Banking. ***ECRIME 2021*** *– SYMPOSIUM ON ELECTRONIC CRIME RESEARCH*

PROJECTS

IDENTIFYING TANDEM REPEATS IN THE DOLPHIN GENOME Genome Alignment, Big data, Python
Performed a proof of concept study to show that DNA fingerprinting via Tandem Repeats can be conducted on bottlenose dolphins. Used tools like Tandem Repeat Finder, minimap2 to calculate and align the TRs of two different dolphins and searching for polymorphic TR sites.

CROP TYPE MAPPING IN KENYA Python, ML, Sequence models
Performed a comparative analysis of Machine Learning and time series algorithms for the task of predicting the crop type of a region in Kenya using Sentinel-2 satellite images. It achieved 65.33 % accuracy using Gradient Boosting Classifier.

CANCER CLASSIFICATION FOR PERSONALIZED MEDICINE PCA, NLP, ML, Python
Built an ML Classifier to categorize cancer into defined classes using genetic mutation text data for personalized medicine. Used TF-IDF and Random Forest classifier (accuracy: 62%).