**Konstantina Kanta**

Phone 004915510314951 ● [konstantinakanta@pm.me](mailto:konstantinakanta@pm.me) ● Hamburg, Germany

**PROFESSIONAL SUMMARY**

Engineer with 5 years of experience in data science, specializing in healthcare data, time series analysis, machine learning, and statistical modeling. Expertise in various areas including predictive modeling, NLP, classification and clustering models, and information retrieval. Highly proficient in Python and *SQL*, with a track record of applying data-driven solutions to complex challenges.

**SKILLS**

**Programming Languages**: Python, R, SQL, C/C++, Java, JavaScript, MATLAB, Perl, HTML, PHP, CSS

**Data Tools & Platforms**: Jupyter Notebook, MongoDB, Apache Spark, Hadoop, Databricks, Vega-Lite, Weka, Datarobot

**Cloud & DevOps**: Microsoft Azure, Kubernetes, Docker, Openstack

**Version Control & Collaboration**: Git, JIRA

**Operating Systems**: Ubuntu, Linux, Windows, MacOS

**Software & IDEs**: Autocad, Android Studio, Arduino IDE, VSCode, Excel, Microsoft Office, LATEX

**Visualization & Analytics**: PowerBI, Tableau, Vega-Lite

**EXPERIENCE**

**Research Scientist** September 2022 – Present

Beiersdorf AG (Nivea) | Hamburg, DE

* Led the Female Health project, where I developed and implemented time series forecasting and clustering models to predict patterns and correlations between women’s menstrual cycles and skin health.

Applied Methods: Data Preprocessing | EDA | Outlier Detection & Handling | Standardization | K-Means Clustering

* Supported the development of diffusion models on selfies and high quality skin images to generate high-accuracy zoomed-in skin images from selfie inputs, that ensure high resolution and precision for future predictions with minimal input data.

Applied Methods: Diffusion Models | U-Net | Embedding-based conditioning | Hugging Face Diffuser

* Developed and deployed a time-series machine learning model that identifies stable phases in skin moisture levels using differential-based feature extraction from sensor data, retaining only the most reliable points to enhance the final accuracy and optimised the speed performance.

Applied Methods: Time Series | Data Filtering & Signal Smoothing | Model Optimization | Performance Tuning

* Statistical analysis using multivariate regression, with confounder control through covariate adjustment, interaction terms, and variance inflation factor assessment to estimate the impact of external factors on a person’s age.

Applied Methods: Statistical Modelling |Ordinary least squares (OLS) regression | Bootstrapping

* Supervised and mentored working students and new incomers.
* Represented the team and projects within different settings, including presentations for members of the European Parliament.

**Data Scientist** February 2022 – August 2022

Shutterstock, Inc. | Dublin, IE

* Led the Search Fusion project to merge search results from various Shutterstock recommendation systems based on rank.
* Implemented reciprocal rank fusion, achieving a 10% improvement in accuracy over benchmark systems.

**Cloud Integration Engineer** June 2020 - February 2022

Ericsson | Dublin, IE

* Acted as a primary support contact for internal stakeholders across multiple global teams, handling numerous Customer Support Requests (CSRs), providing technical assistance, guidance and communication, offering a high level of customer service and strengthening my experience in client-focused support roles.
* Led the integration and deployment of Ericsson cloud solutions, troubleshooting and tracing with technical decision making, along with technical knowledge in areas such as Linux, MySQL, Kubernetes, Docker, and Openstack.
* Integrated Cloud Container Distribution (CCD) clusters, implementing and running applications from Docker images organized in separate pods using Kubernetes.
* Responsible for integration planning and deployment of Ericsson’s Cloud Execution Environment, handling test cases as part of acceptance testing for Software Defined Infrastructure (SDI), and supporting international Ericsson’s customers..
* Additional experience includes designing Low Level Design (LLD), VNF, SDN, Linux, and Openstack.
* Some of the customers I worked for included AT&T Inc., Three IE, NTT Docomo Inc., and Bharti Airtel.

**Database and Information Systems Demonstrator**  Sept. 2019 - Dec. 2019

University College Dublin | Dublin, IE

* + Managing the correction and the grading of students’ final SQL project assignments.
  + Demonstrator of the undergrad module for second year students of UCD Computer Science.
  + Responsible for the correction and the grading of student’s weekly SQL assignments, as well as final project assignment.

**EDUCATION**

**University College Dublin** | Master of Science in Computer Science Sept. 2019 - Aug. 2020

* GPA: 3.57/4.2 (2:1 Honours) 90 ECTS
* Advance Machine Learning, Big Data, NLP, Multi-agent Systems (AI), *SQL* Programming, Data Visualisation.

**Democritus University of Thrace** | Master of Science in Electrical and Computer Engineering Oct. 2013 - Nov. 2018

* GPA: 7.8/10.0 (2:1 Honours) 300 ECTS
* Machine Learning, Robotics, Numerical Analysis, Data Mining, Information Retrieval, Linear Algebra.

**ACADEMIC PROJECTS**

**Fusing Rankings in Information Retrieval with Voting Systems, Diploma Thesis, DUTh**

* Comparative analysis and evaluation of different voting systems used for a given task in Information Retrieval.
* The study was implemented in Python with the use of pandas, numpy, etc and evaluated with Trec Eval.

**Implementation of my own multi-label classifier from scratch, University College Dublin**

Implementation of the binary relevance algorithm (with the option that uses under-sampling) and classifier chains from scratch.

**Comparison of the performance of the three different models.**

Deep learning in chest x-ray images for Pneumonia Detection, University College Dublin

A build of a logistic regression (LeNet-5 algorithm) model, a convolutional neural network model, coupled with data augmentation techniques and a VGG16 model to detect pneumonia in chest x-ray images. Comparison of the performance of the four different models, implemented in Python with the use of scikit learn, openCV, keras, pytorch.

**Data Mining academic projects, University College Dublin**

* Classification using Decision trees, Nearest neighbour algorithms, Naive Bayes and Regression analysis
* Clustering using K-means, Neural Networks, Hierarchical Clustering, DBSCAN
* Association rules, Data pre-processing, Data warehouse, Ensemble learning, Dimension reduction

**VOLUNTEERING**

**Information Technology Consultant** Dec 2016 - Oct 2017

Electrical and Computer Engineering Student Conference of Greece | Xanthi, Greece

* Website Developer for ECESCON 10, 2017 (Panhellenic Electrical and Computer Engineering Students Conference): Front End Developer.
* Found at: http:/www.sfhmmy.gr