**7PAM2002 Data Science Project Module**

**Choosing a Project Form**

**Semester B 2024/2025**

This form will be used by your supervisor to agree your project topic and dataset. Complete as much of the form as possible then submit the form into ‘Assignments’ on the Project Module Canvas site.

**Student Full Name:** NANDANA VIJAYAN

**Student ID:** 23038331

**Email Address:** [**nvnair2@gmail.com**](mailto:nvnair2@gmail.com)or[**nv23abd@herts.ac.uk**](mailto:nv23abd@herts.ac.uk)

**Course:** 1 year

**Semester intake to the course:** A

|  |  |
| --- | --- |
| SECTION A |  |
| **What is your proposed project title or topic?** | Evaluating the effectiveness of SMOTE-ENC in Class Imbalance Medical Data. |
| **What is the Research Question for your project?** | How does the SMOTE-ENC algorithm perform in balancing datasets with both nominal and continuous features for COPD classification using the ExaSens dataset? |
| **Dataset website address (or organisation and person the dataset is from).** | <https://ieee-dataport.org/open-access/exasens-novel-dataset-classification-saliva-samples-copd-patients> |
| **Where was the data originally collected? (who, when ,where)** | This repository introduces a novel dataset for the classification of Chronic Obstructive Pulmonary Disease (COPD) patients and Healthy Controls. The Exasens dataset includes demographic information on 4 groups of saliva samples (COPD-HC-Asthma- Infected) collected in the frame of a joint research project, Exasens (https://www.leibniz-healthtech.de/en/research/projects/bmbf-project-exasens), at the Research Center Borstel, BioMaterialBank Nord (Borstel, Germany). The sampling procedure of the patient materials was approved by the local ethics committee of the University of Luebeck under the approval number AZ-16-167 and a written informed consent was obtained from all subjects. A permittivity biosensor, developed at IHP Microelectronics (Frankfurt Oder, Germany), was used for the dielectric characterization of the saliva samples for classification purposes (https://www.mdpi.com/2227-9032/7/1/11). |
| **What type of data are you using? (e.g. image/ tabular/category/continuous etc)** | Tabular |
| SECTION B |  |
| Can you attend all 6 supervision sessions on-campus? If not state reason. | Yes |
| What was the subject of your BSc degree and any other Master’s degrees you have taken? | I have completed a Bachelor of Computer Applications (BCA) and a Master of Computer Applications (MCA) |
| What are your career aims and/or the industry sector you would like to get into? | I aim to build a career in the field of data science and machine learning, contributing to innovative solutions in industries like healthcare, technology, or finance. |
| A brief account of your programming/data science experience (including work or placement). | I have hands-on experience in programming and data-science through academic projects and practical implementations. During MCA, I was a part of group project and completed a mini project titled Criminal Detection: Study of Activation Functions and Optimizers in Model Building, which explored variations in model accuracy based on activation functions and optimizers.  In my BCA, our group developed Jansevak, a website designed to facilitate online charity in education and healthcare fields. Additionally, I worked on a Grocery Store Management App during my MCA, an application for managing grocery store operations online.  Also, I worked as a software engineer for 2.5 years in the Laboratory Information System (LIS)domain . My role involved leveraging my skills as an LIS Analyst and Interface Engineer to develop and optimize solutions for efficient laboratory operations. |
| Any work experience (including non-computing related). | I worked as a Software Engineer for two and a half years at Amrita Technologies, Kochi,Kerala,India. |
| Do you have any hobbies, activities or interests. | I enjoy coding, exploring new technologies, solving puzzles, reading and listening to music in my free time. |