**PBL: Application of Inferential Statistics and Information Theory to a Computer Science Problem**

**Objective:**To select a relevant computer science problem, apply the core principles of inferential statistics and information theory to analyze it, build predictive models, and derive scientifically sound conclusions and innovative recommendations.

**Example domains include:**

* **Network Security**: Intrusion Detection, Spam Filtering
* **Human-Computer Interaction**: User Behavior Analytics, A/B Testing
* **Systems Performance**: Hardware Reliability Forecasting, Load Prediction
* **Natural Language Processing**: Document Classification, Topic Modelling
* **Bioinformatics**: Gene Sequence Analysis, Disease Prediction

**Design:** Identify a computer science project as a use case for Inferential Statistics & Information theory and collect the required data and information.

**Example of project:** **Predicting and Analyzing Network Intrusion Using Inferential Statistics and Information Theory**

**Implement: -** Formulate the statistical steps: data exploration, modeling, estimation, inference, and prediction/forecasting, with the help of python/R.

**-** Implement discussed algorithms (at least 2 algorithms)

**Evaluate:** Interpret results and suggest improvements, to promote new ideas and innovative solutions.

**Based on feedbacksprovide a compressed folder:**

* **Final report**
* **Data (Excel or CSV)**
* **Code (Python or R)**
* **PPT for final presentation**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Component** | **Milestone I** | | Fail to meet expectations | Meet expectation | Excellent |
| * **C1** | * Identify a computer science project (Your choice needs to be linked with the course objectives (ILO1, ILO2, ILO3, ILO4, ILO5). | |  |  |  |
| * **C2** | * Explain how this project is an appropriate use case for data analytics, statistical analysis and prediction/forecasting | |  |  |  |
| * **C3** | * Collect the required data and information | Describe the collection process (survey, website, etc..) |  |  |  |
| Provide metadata. |  |  |  |
| Provide a description of anonymization and data protection procedures. |  |  |  |
| Examine reliability, consistency and validity of sources |  |  |  |
| In the case of sampling procedures   * Describe the used sampling procedure. * Evaluate how the sample relates to the population of interest.   In the case of a survey   * Provide used literature for survey development * Mention nonresponses bias. * Assess sources of “validity evidence”. |  |  |  |