Proforma for Submitting a Proposal to PGPEC for Introducing a New Subject

1. Department Proposing the Subject : Computer Science and Engineering

2. Subject Number : To be decided

3. Subject Title : Data Analytics

4. Subject Structure and Credit : 3-0-0 Credits: 3

5. Status of the Subject : Elective at undergraduate level

6. Pre-requisites for the Subject : a. Programming and data structures

b. Probability and statistics

7. Objective and Content Objective : Introduction to theoretical techniques and practical tools used in data analytics. Applications in various engineering and scientific domains.

Content (in 100-150 words): The course will cover fundamental algorithms and techniques used in data analytics. The statistical foundations will be first covered, followed by various machine learning and data mining algorithms used in data analytics. Technological aspects like data management, scalable computation, and visualization will also be covered. The course will provide exposure to theory as well as practical systems and software used in data analytics. As a part of the course students will take up analytics project from various domains.

(See annexure for the details syllabus of the course.)

8. Name of the Faculty Members of the Department who have Necessary Expertise and will be Willing to Teach the Subject Currently: : a) Sudeshna Sarkar b) Pabitra Mitra

9. **Will the Subject Require Appointment of Adjunct Faculty**? If yes the Number of Such Adjunct Faculty: NO

10. **Do the Content of the Subject Have an Overlap** with any Other Subject Offered in the Institute? If yes give Details

a) Number and the name of the existing subject : Two (Machine learning, Data mining)

b) Approximate percentage of Overlap : 30%

11. **Suggested Reading**: Text Books/Reference Books: See Annexure for Syllabus and Text/Reference Book

12. **The Name of the Departments/Centers** Whose Students are Expected to take up the Course:

All departments

Date: **(Signature of the Head of the Department)**

ANNEXURE

CS Data Analytics 3-0-0 Credits: 3

Syllabus:

Data management and indexing

Data representation and characterization

Basic statistical analysis tools and models

Data analytics programming languages – R, SPSS, Matlab, Python

Association and correlation analysis- regression models

Predictive analytics

Exploratory analysis

Feature engineering

Visualization

Scalable and parallel computing

Text analytics

Case study and project

References:

Text Books:

1. An Introduction to Statistical Learning: with Applications in R, G James, D. Witten, T Hastie, and R. Tibshirani, Springer, 2013

2. Software for Data Analysis: Programming with R (Statistics and Computing), John M. Chambers, Springer

3. Mining Massive Data Sets, A. Rajaraman and J. Ullman, Cambridge University Press, 2012