Indian Institute of Technology, Kanpur

Proposal for a New Course

1. Course No: **EE XXX** 708
2. Course Title: **Fundamentals of Data Science and Machine Intelligence**
3. Per Week Lectures: **3** (L), Tutorial: **0** (T), Laboratory: **0** (P), Additional Hours[0-2]: **0** (A), Credits (3*L): **09** Duration of Course: **Full Semester**
4. Proposing Department/IDP : **Electrical Engineering (EE)**
   Other Departments/IDPs which may be interested in the proposed course: **All Departments**
   Other faculty members interested in teaching the proposed course: **Dr. Ketan Rajawat**
5. Proposing Instructor(s): **Dr. Rajesh M. Hegde**
6. Course Description:

A) Objectives:

This course aims to equip students with the fundamental concepts and techniques central to the fields of exploratory data analysis, statistical inference, and machine learning leading to machine intelligence. Students from all disciplines, both engineering and sciences can develop proficiency in data analysis/visualization, statistical data analysis, machine learning algorithms, and machine learning tools, enabling them to obtain actionable insights from complex datasets in various domains by completing this course. Students will be exposed to the design and implementation of machine learning models, and handling AI frameworks in Python via a course project. By the end of the course, students are expected to be able to design, implement, and evaluate machine intelligence in general and thus preparing them for careers in data science, artificial intelligence, and related disciplines. **The course is targeted at all engineering and science disciplines who wish to understand the emerging and popular paradigm of Data Science and Machine Intelligence.**

B) Contents: (40 lectures, each of 50 minutes duration)

S No.	Broad Title	Topics	No. of lectures
1	Foundational principles of data science and machine intelligence	1. Introduction to Data Science (DS), and Machine Intelligence (MI) 2. Introduction to Algorithms, Models, Optimization Techniques for DS/MI. 3. Introduction to Supervised Learning, Unsupervised Learning, and other learning techniques	3
2	Statistical data analysis, visualization, and inference	1. Statistical Data Analysis: Descriptive Statistics, Exploratory Data Analysis (EDA), Hypothesis Testing, Correlation and Covariance 2. Data Visualization: Histogram, Scatter Plot, Box Plot and other plots 3. Model Visualization: Confusion Matrix, ROC curve Inference: Point Estimation, Confidence intervals, Bayesian Inference	6
3	Regression analysis and modeling	1. Linear Regression Modeling 2. Non-Linear Regression Modeling 3. Logistic Regression 4. The Bias-Variance Decomposition 5. Bayesian Linear Regression	6

4	Clustering, Decision Trees, PCA, ICA, Vector Quantization	1. Distance Measures 2. Clustering: K-Means, clustering variants, Hierarchical clustering 3. Decision Tree: Tree construction, Pruning 4. Random Forests: Ensemble learning, Out-of-bag error, Feature Importance 5. Bosting/Bagging	9
5	Gaussian Mixture modeling	1. Probabilistic distance measures 2. Univariate and Multivariate distributions 3. EM algorithm 4. Gaussian Mixture modeling and applications in machine intelligence	3
6	Artificial Neural Networks	1. Feed Forward neural networks 2. Network training 3. Error backpropagation, regularization.	2
7	Deep Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks, and Auto encoders	1. Deep feed forward networks, Regularization and optimization for training deep models 2. Convolutional Neural Networks: Convolution operation, convolution and pooling, variants of convolution function. 3. Recurrent Neural Networks 4. Auto encoders and Applications	9
8	Applications of machine intelligence in science and engineering	1. Case studies of data science/machine intelligence in pure sciences 2. Case Studies of data science/machine intelligence in engineering	2

C) Recommended pre-requisites :

Undergraduate/Graduate Mathematics, Calculus, Linear Algebra, Probability, Statistics and Programming

D) Short summary for including in the Courses of Study Booklet:

The course on Data Science and Machine Intelligence aims to equip students with a comprehensive understanding of the foundational principles and advanced techniques in machine learning/AI. The course will blend theory with practical issues such as AI frameworks and programming. The course will provide a detailed understanding of exploratory statistical data analysis, visualization, and inference. Regression analysis and modeling, Classification Modeling, Decision Trees and Random Forests, Boosting/Bagging, Clustering, LDA, PCA, vector quantization, Gaussian Mixture modeling, Artificial Neural Networks, Deep Neural Networks, Convolutional Neural Networks, and Recurrent Neural Networks. Applications of machine intelligence covering areas in both science and engineering will be discussed using data spread across disciplines and applications. This course will have significant focus on regular hands on assignments and end with a course mini project to be implemented in Python using AI/ML frameworks.

7. Recommended Books (Reference Books) :

1) Alpaydin, E., ***Introduction to Machine Learning***. Cambridge, MA: MIT Press. ISBN: 9780262043793, Mar. 2020
2) Duda and Hart, ***Pattern Classification***, Wiley, 2001
3) C. Bishop, ***Pattern Recognition and Machine Learning***, Springer, 2006.
4) Goodfellow, I.; Bengio, Y. & Courville, A., ***Deep Learning***, MIT Press, 2016.
5) S. Theodoridis and K. Koutroumbas, ***Pattern Recognition*** Second Edn, Elsivier, 2003
6) B. Yegnanarayana, ***Artificial Neural Networks***, PHI, 1999.
7) Simon Haykin, ***Neural Networks and Learning Machines***, Pearson, 1999.
8) Joel Grus, ***Data Science from Scratch: First Principles with Python***, O'Reilly Media, Inc 2019
9) Andreas C. Müller, Sarah Guido, ***Introduction to machine learning with Python : a guide for data scientists,*** O'Reilly Media, Inc., 2016
10) Andriy Burkov, ***The Hundred-Page Machine Learning Book***, 2019
11) Jake VanderPlas, ***Python Data Science Handbook: Essential Tools for Working with Data***, O'Reilly Media, Inc 2016
12) Wes McKinney, ***Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython***, O'Reilly Media, Inc 2017
13) Aurélien Géron, ***Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems***, O'Reilly Media, Inc 2017
14) Thomas Nield, ***Essential Math for Data Science***, O'Reilly Media, Inc., 2022
15) Peter Bruce and Andrew Bruce, ***Practical Statistics for Data Scientists***, O'Reilly Media, Inc, 2017
16) John D. Kelleher, Brian Mac Namee, Aoife D'Arcy, ***Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies***, MIT Press, 2015
17) John Hearty, ***Advanced Machine Learning with Python***, Packt Publishing, 2016
18) Kevin Patrick Murphy, ***Probabilistic Machine Learning: Advanced Topics***, MIT Press, 2023

8. Any other remarks: **Hands on assignments and a mini project to be implemented in Python/AI Frameworks will form an important part of the grading scheme for this course.**

Dated: **11/08/24** Proposers: **Dr. Rajesh M. Hegde**

DPGC Convener (EE) : _____

The course is approved / not approved

Chairman, SUGC/SPGC

Dated: _____