Course Number	EP439
Course credit (L-T-P-C)	2-0-2-6
Course title	Data Science for Physics
Learning mode	Offline
Learning objectives	• An introduction to data science career path for physicists. • Understanding the basics of machine learning and ML model building. • Exposition to popular python-based environments like Jupyter, Kaggle which are used industry-wide for AI/ML or data science applications. • Using state-of-the-art libraries like pandas and sklearn to preprocess the data, apply ML models, validate, and test predictions. • Hands-on experience through real-world projects.
Course description	Data science is increasingly becoming an essential skill for physicists. While there are numerous courses and programs on data science offered across various media, these are almost invariably targeted at computer science graduates and industry professionals. This course is designed to bridge this gap by introducing essential data science techniques from the perspective of applications in physics research and prepare learners for advanced courses in ML/AI/Data science.
Course content	Python and programming environments: review of python, setting up local python development environment, setting up Jupyter, introduction to Kaggle-cloud based python notebook and data science platform, creating first python program on Kaggle/Jupyter notebook, basics of numpy library, file versioning using github. The what and why of machine learning, mathematical basis of ML – setting up a problem, example of linear and polynomial regression; statistical learning theory – bias, variance, model complexity; cost function, gradient descent, basics of supervised and unsupervised learning, regression with multiple variables, feature normalization, basics of neural networks, building first ML model – handling data for training, testing, and validation, types of models, using scikit-learn library, ML pipelines; data science techniques – pandas, data cleaning, data visualization. Hands-on project – detection of gravitational waves – introduction to gravitational waves, Fourier transform, noise, GW signal analysis, data fitting. Hands-on project – Detection of signal and background events at the LHC and other colliders.
Pre-requisites	• Linear algebra, matrices, vector algebra • Basic familiarity with programming in Python
Learning outcomes	Upon successful completion of this course, students will be able to: • write intermediate-level programs in Python, define functions, import and use libraries. • Work on projects in Jupyter environment, and collaborate on group projects on platforms like Kaggle, and github. • Understand the fundamental concepts of machine learning and theoretical understanding of how ML models are developed. • Understand and manipulate data for training, validating, and testing predictions of ML models. • Use various python libraries like scikit-learn, pandas, numpy, etc. to create ML pipelines that take in given data and generate predictions. • Get exposure to real-world usage of data science techniques in trending research areas.
Assessment method	Group project (P), Assignments (A), MidSem (MS), EndSem (ES). Internal (P+A)=40%, MS=30%, ES=30%
Textbooks and references	Main references: • Christopher Bishop, Pattern Recognition and Machine Learning, Springer, 2007 • Introduction to Machine Learning Edition 2, by Ethem Alpaydin • Machine Learning. Tom Mitchell. First Edition, McGraw- Hill, 1997. Additional References: • A high-bias, low-variance introduction to Machine Learning for physicists, Pankaj Mehta,

	Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, David J. Schwab, 2019, Phys. Rep. 810, 1. • John Hopcroft, Ravindran Kannan, Foundations of Data Science, 2014. • I. Goodfellow, Y. Bengio, A. Courville. Deep Learning. MIT Press, 2016. • Machine learning & artificial intelligence in the quantum domain, Vedran Dunjko, Hans J. Briegel, arXiv:1709.02779 • Andrew Ng's lectures on machine learning, Coursera, https://www.coursera.org/learn/machine-learning-course/