Course Title	**Data Analysis and Modeling**
Course Code	**EES 211 2.0**
Credit Value	02
Status	Core
Year / Level	Year 2
Semester	1
Theory: Practical: Independent Learning	30: 00: 70
Other: Pre-requisite Course/s	EES 112 2.0 Probability and Statistics

Aim of the Course:

To provide students with a working knowledge of statistical methods of data representation, analysis, modeling and interpretation and to help students develop skills in using a computer software for data analysis and data modeling. In addition, to help students gain an understanding of how statistics can be used for applications in electronic based development projects.

Intended Learning Outcomes:

On the successful completion of this course, the student should be able to:

Use a statistics software package to represent data using suitable data structures and types.

1. Perform basic arithmetic and statistical operations such as computing the mean, median, variance and standard deviation of a given data set, both manually and through the use of statistical software.
2. Compute the line of best fit for a given data set using regression analysis and plot the line with the aid of software. Using the regression fit, compute the coefficient of variation and coefficient of determination and interpret the values obtained for a given data set.
3. Obtain the residual plot for a data set fitted with a linear regression model and suggest how the model can be optimized. In addition explain how this method can be extended into multiple regression analysis.
4. Graphically represent a given data set and estimate the most suitable curve which describes the data pattern using polynomial curves of varying degrees, exponential and logarithmic curves and obtain the equation of the curve of best fit.
5. Model data using common statistical distributions such as normal distribution, poisson distribution and binomial distribution and obtain useful estimates for the given problem.
6. Use moving average method, triangular smoothing and Savitzky-Golay filters to smooth a given data set which represent a physical system.
7. Define and explain the categories of machine learning, and briefly explain the process of machine learning.

Course Content:

Introduction to data analysis and modeling: Introduction to the statistical software, Importing and preparation of data using software, Handling Data, Data types, Operators and Data Structures, Functions,

Basic Statistical Plots, Handling errors in software. Introduction to Linear regression analysis: The least squares method, Goodness of fit, Residual values, Fit optimization methods, Coefficient of correlation & coefficient of determination, Linear regression analysis using software, Introduction to multiple regression analysis. Introduction to curve fitting: Curve fitting using software, Polynomial fitting, Exponential and logarithmic curve fitting, Extrapolation, Inference in curve fitting. Data visualization: Statistical distributions using software. Introduction to signal smoothing: Moving average, Triangular smoothing, Savitzky-Golay smoothing. Inferential statistics: Introduction to machine learning, Multivariate plots & clustering.

Scope and Schedule of Teaching - Learning Activities:

Topic No.	Topic / Sub Topic	No. of Hrs.			Teaching Method	Assessment Criteria	ILO Alignment
		T	P	IL			
1	Introduction to programming and data analysis with R Data structures, Basics of programming, Data import and export, Data wrangling	4	0	8	Lecture / Practical	Assignment (5%) In-class Assignment (10%)	1,2
2	Data visualization with R	4	0	6	Lecture/ Practical		3
3	Linear regression analysis: Simple linear regression, Multiple linear regression, Multicollinearity, Model adequacy checking, Variable selection	6	0	10	Lecture/ Practical		3,4
4	Nonlinear regression: Polynomial regression, Regression splines, Smoothing splines	4	0	10	Lecture / Handout 3 /Tutorial	Assignment (25%)	5
5	Introduction to machine learning	2	0	12	Lecture / Handout 4/ Tutorial		5
6	Supervised learning algorithms: Tree-based methods, Support vector machines	6	0	14	Lecture / Tutorial		6
8	Unsupervised learning algorithms: Principal component analysis, Clustering methods	4	0	10	Lecture		7
	Total	***30***	***00***	***70***			

Linking Program Outcomes with ILOs:

Program Outcomes:

1. Demonstrate competency in theoretical knowledge and practical and/or technical skills in the respective field of specialization.
2. Communicate efficiently and effectively in the respective field of specialization using written, oral, visual and/or electronic forms.
3. Facilitate and participate as an empathetic and emotionally intelligent team player with leadership qualities, in a group, diverse team or organization.
4. Apply subject-specific knowledge and skills creatively to solve real-world problems by making context-specific operational decisions while adapting to changing environments.

5. Integrate creativity, innovation, and entrepreneurial and managerial proficiencies to build values.
6. Implement subject-based solutions in keeping with ethical, societal and environmental norms and need for sustainable development.
7. Secure lifegoals through lifelong learning with the aim of scholarly advancement and/or strengthening professional skills, and ensuring the betterment of the community.

	PLO 1	PLO 2	PLO 3	PLO 4	PLO 5	PLO 6	PLO 7
ILO 1	***	**		**			
ILO 2	***	**		***		*	*
ILO 3	***	**		***		*	*
ILO 4	***	***		***		*	**
ILO 5	***	**		***		*	**
ILO 6	***	**		**		*	**
ILO 7	***	**		*		*	**

*** - *Strongly Linked;* ** - *Medium linked;* * *Weakly linked*

Mode of Assessment:

Formative Assessment (FA): FA1 5% + FA2 10% + FA3 25% = 40% of Total Marks

Summative Assessment (SA): End Semester Examination: 1-hour paper with 2 Structured questions to develop computational solutions (60%) of Total Marks

References

- Introduction to Linear Regression Analysis - D.C. Montgomery, E.A. Peck, G.G. Vining, John Wiley & Sons, 4th edition, 2012
- Data Fitting and Uncertainty, Tilo Strutz, Springer Vieweg, Edition 2 , 2016
- Introduction to Statistics and Data Analysis for Physicists, Bohm G., Gunter Z.,