**24SF805 Data Science for Social Studies 2 0 1 3**

The "Data Science for Social Studies" course is designed to equip students with essential data science skills tailored for social science applications. This course aims to bridge the gap between data science and social studies, enabling students to harness the power of data to address real-world social issues and inform policy decisions.

## Learning Objectives

**LO1** Understand the fundamental concepts of data science and its applications in social studies.

**LO2** Develop skills in data collection, cleaning, and preprocessing specifically for social science data.

**LO3** Gain proficiency in using data science tools for analysis, visualization, and interpretation of social science data

**LO4** Apply data science techniques to address real-world social issues and policy questions.

## Course Outcomes

**CO1** Ability to apply data preprocessing techniques to prepare social science data for analysis.

**CO2** Ability to perform exploratory and inferential data analysis using statistical and machine learning methods.

**CO3** Apply data science techniques to address real-world social issues and policy questions.

## Course Contents

**Module 1 - Data collection, cleaning and preprocessing:** Types of data in social studies: qualitative vs. quantitative data. Sources of social science data: surveys, censuses, administrative data, social media, and open data sources. Data cleaning techniques: handling missing data, detecting and treating outliers, data normalization.

**Module 2 - Statistical Methods for Social Data**: Descriptive statistics: measures of central tendency and dispersion. Correlation and causation: understanding relationships between variables. Probability Distributions. Hypothesis testing: t-tests, chi-square tests, and ANOVA.

**Module 3 - Machine Learning for Social Studies:** Supervised learning: regression and classification techniques (e.g., linear regression, logistic regression, decision trees), Unsupervised learning: clustering (e.g., k-means, hierarchical clustering) and dimensionality reduction (e.g., PCA). Model evaluation and validation: accuracy, precision, recall, and cross-validation

**Module 4 – Making Impact with Analysis**: Principles of effective data visualisation: storytelling and reporting using dashboards. Case studies in applying data driven techniques to inform policy. Individual student presentations on applying data science concepts to research data collected till date.

**Evaluation and Assessment**

1. Quizzes & Assignments
2. Lab assessment of applied concepts
3. Case studies presentation

**Evaluation Pattern**

| **Category** | **Marks** |
|---|---|
| Continuous Assessment | 20 |
| Mid-Term | 30 |
| End Semester | 50 |
| **Total** | **100** |

**Reference Material**


1. Healey, J.P. Statistics: A Tool for Social Research. 8th Edition. Belmont, CA: Wadsworth. 2009.
2. Agresti, A. (2018). Statistical methods for the social sciences. Boston: Pearson
3. Case Studies from "Turn Data into Policy": https://www.datatopolicy.org/use-cases#Discover
4. R for Non-Programmers: A Guide for Social Scientists, Online Book available at: https://bookdown.org/daniel_dauber_io/r4np_book/
5. James, Witten, Hastie, Tibshirani, An Introduction to Statistical Learning with Applications in R, Online link: https://hastie.su.domains/ISLR2/ISLRv2_website.pdf
6. Intro to R for Social Scientists, Jasper Tjaden, Youtube playlist available online at: https://www.youtube.com/playlist?list=PLr43hk2e3hFMg4tZdJsN0qzG5YkQB3A1c