|  |  |
| --- | --- |
| **DEBRE BERHAN UNIVERSITY COLLEGE OF COMPUTING DEPARTMENT OF Data Science**  **Course Syllabus** | |
| Program | B.Sc. in Data Science |
| Name of Course: | **Machine Learning** |
| Course Code: | DaSc3031 |
| Year/Semester | Year 3, Semester I |
| CP/ECTS | 5 |
| Study Hours: | 30 Hours (Lecture, F2F), 60 Hours (Lab) |
| Independent Hours | 30 Hours |
| Assessment hours | 15 Hours |
| Total Notional Hours | 135 |
| Pre-requisite | Artificial Intelligence (CoSc2122) and  Regression Analysis for data Science(Stat2052) |
| Name(s) of academic staff: | Wudneh K., Bestlot Y, Beyenech A. |
| **Summary:** This course covers Machine Learning theory, algorithms, and applications. Machine Learning is currently at the heart of Artificial Intelligence. The course covers a number of machine learning methods with a focus on prediction. Topics include core methods such as supervised learning (regression and classification), unsupervised learning (clustering and association), principal component analysis, Bayesian estimation, neural networks; common practices in data pre-processing, hyper-parameter tuning, and model evaluation. The course will be project-oriented, with emphasis placed on writing software implementations of learning, Algorithms applied to real-world problems. | |

1. **Course Learning Outcomes (CLO):**

*At the end of the course, the students will be able to:*

**CLO1:** Describe the fundamental concepts and types of machine learning ML and Data mining techniques and algorithms.

**CLO2:** Perform data preprocess, data cleaning and feature engineering and apply machine learning methods to achieve a learning goal within an intelligent system.

**CLO3:** Implement and utilize popular libraries and frameworks for machine learning.

**CLO4:** Choose and apply machine learning techniques to real-world problems in various domains, such as healthcare, finance, and marketing and develop models.

**CLO5:** Evaluate the performance of machine learning models and judge the suitability of a machine learning paradigm for a given problem and the available data.

**CLO6:** Interpret and communicate the results of machine learning models to stakeholders in a clear and understandable manner.

**CLO7:** Recognize ethical considerations and potential biases in machine learning algorithms, and apply fairness and transparency principles in model development and deployment.

**CLO8:** Stay updated with the latest advancements in machine learning research and techniques, and continue to learn and improve skills in this rapidly evolving field.

1. **Mapping CLO to PO and Assessment Methods**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CLO** | **Program Learning Outcomes (PLO)** | | | | | | | | | **Assessment Methods** | |
| **PLO1** | **PLO2** | **PLO3** | **PLO4** | **PLO5** | **PLO6** | **CLO8** | **PLO9** |  | |
| CLO1 | X |  |  |  |  |  |  |  | Lab/Quiz, Mid-Exam, Final Exam | |
| CLO2 | X |  |  |  |  |  |  |  | Lab/Quiz, Mid-Exam, Mini Project, Final Exam | |
| CLO3 |  |  |  |  | x |  |  |  | Lab, Mini project and Final | |
| CLO4 |  |  | X | x |  |  |  |  | Quiz/Lab, Mini project Final | |
| CLO5 |  | X |  | x |  |  |  |  | Mini project, Final Exam | |
| CLO6 |  |  |  |  |  |  | x |  | Mini project, Report | |
| CLO7 |  |  |  |  |  | x |  |  | Mini project, Final Exam | |
| CLO8 |  |  |  |  |  |  |  | x | Mini project, Final Exam | |

**Teaching/Learning Strategy:** The main teaching method will be Lecture, case study, Laboratory, project work and cooperative learning supported by tutorial sessions. The students will be expected to perform presentations & self-learning exercises on different machine learning techniques.

1. **Course Content and CLO Mapping and Teaching Methods**

|  |  |  |
| --- | --- | --- |
| **Course Content** | **CLO** | **Teaching**  **Methods** |
| **Chapter 1: Introduction to machine learning**  Introduction to machine learning concepts:History of Machine Learning, Data Mining (goals, process and techniques), and Knowledge Discovery in Databases (KDD).  Data processing, Data cleaning and transforming Training and Testing Machine Learning models, Evaluation metrics, Components of Learning, Machine learning models.  Why Machine Learning, Data and Data Preprocessing and types of machine learning techniques. Role of AI in Machine Learning.  **Lab Content***:**Data Preprocessing case study.* | CLO1  CLO2 | Lecture |
| **Chapter 2: Supervised learning**  **Regression:** Simple linear regression, Multiple Linear Regression, Polynomial Regression, Random Forest Regression, [Regularization](https://www.geeksforgeeks.org/regularization-in-machine-learning/)(Ridge Regression, Lasso Regression), Gradient Boosting Regression and Regression Models Performance.  **Classification:** Decision Tree, K-Nearest Neighbors (K-NN), Logistics regression, Naïve Bayes, Support Vector Machine (SVM) and Kernel SVM, Bias-variance tradeoff, discriminant Analysis, Classification Models Performance.  **Lab Content***:**Implementation of regression and classification algorithms using real world scenarios.* | CLO2,  CLO3,  CLO4,  CLO5 | Lecture, Laboratory work |
| **Chapter 3: Unsupervised learning**  **Clustering**: Hierarchical clustering, k-means, C-means, mixture models, Neural Networks, Anomaly detection, dimension reduction, principal components analysis, independent component analysis, factor analysis and Singular value decomposition (SVD). Nonlinear dimensionality reduction, Gaussian Mixture models and cluster validation.  **Association Analysis:** A prior algorithm, Frequent Pattern Growth Algorithm, Hidden Markov model, Monte Carlo prediction.  *Lab: Apply and evaluate clustering models, Implement Apriori algorithm for Market basket analysis* | CLO2,CLO3,CLO4,CLO5 | Lecture, Laboratory work |
| **Chapter 4: Building and Evaluation Machine Learning Models**  **Dimensionality Reduction Technique:** Linear Discriminant Analysis ,[Principal Component Analysis (PCA)](https://www.geeksforgeeks.org/principal-component-analysis-pca/), [t-distributed, Stochastic Neighbor Embedding (t-SNE)](https://www.geeksforgeeks.org/ml-t-distributed-stochastic-neighbor-embedding-t-sne-algorithm/), [Non-negative Matrix Factorization (NMF)](https://www.geeksforgeeks.org/non-negative-matrix-factorization/), [Independent Component Analysis (ICA)](https://www.geeksforgeeks.org/ml-independent-component-analysis/), [Factor Analysis](https://www.geeksforgeeks.org/introduction-to-factor-analytics/), Nonlinear latent variable models,  Classification Metrics, Accuracy, Precision and recall, ROC curves, Regression Metrics, Cross-Validation. Types of Predictive Models.  ***Lab:*** *Implementing of Model Evaluation and Performance Enhancement techniques* | CLO5, CLO6, | Lecture, Laboratory work , Project work, Report work |
| **Chapter 5: Main Challenges and Real-World Applications of ML**  Data Quality, Over-fitting and Under-fitting, Cross-Validation and Re-sampling methods *(K-Fold Cross-Validation, Bootstrapping,* Jackknife), Gradient descent (batch, stochastic), Bias, bias-variance trade-off, cost function, train and test error.  Performance evaluation methods.  ***Lab:*** *An integrated real-world project for selected domain/area.* | CLO4, CLO7, CLO8 | Lecture, Laboratory, Project & report work, |

**Transferable Skills:** Problem-solving skill, case study analysis, Machine Learning techniques, Collaboration.

1. **Assessment Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No** | **Assessment Task** | **Value/ Mark** | **CLO** |
| 1. | In-class Quizzes and Lab Exams | 15% | 1,2,3, 4 |
| 2. | Mid-Exam | 15% | 1,2,3,4 |
| 4. | Mini Project *(With documentation and report)* | 30% | 2,3,4,5, 6, 7,8 |
| 5. | Final Exam | 40% | 1,2,3,4,5,7,8 |

1. **Text Books and Reference**

**References**

1. Lindholm, A., Wahlström, L., Lindsten, F. och Schön, T. B. (2021). Machine Learning – A First Course for Engineers and Scientists. Cambridge University Press. A draft version is freely available in PDF at: <http://smlbook.org/>.
2. Knox, Steven W. *Machine learning: a concise introduction*. Vol. 285. John Wiley & Sons, 2018.
3. Kuhn, Max, and Kjell Johnson. *Applied predictive modeling*. Vol. 26. New York: Springer, 2013.