III B.Tech I Semester

		L	T	P	C
23A39503	**MACHINE LEARNING & NLP LAB (Professional Core)**	**0**	**0**	**3**	**1.5**

Course Objectives:

1. To provide hands-on experience with fundamental Machine Learning and NLP algorithms.
2. To enable students to preprocess, train, and evaluate models on structured and unstructured data.
3. To develop NLP applications like classification, summarization, and conversational agents.
4. To equip students with the skills to use ML libraries and frameworks for solving real-world language problems.

Course Outcomes:
After successful completion of this lab, students will be able to:

1. Apply supervised and unsupervised ML algorithms on various datasets.
2. Preprocess and analyze textual data for NLP applications.
3. Implement machine learning pipelines using tools like scikit-learn and NLP libraries.
4. Develop and deploy basic NLP applications such as chatbots, sentiment analyzers, and translators.
5. Evaluate models using appropriate metrics and improve them using tuning techniques.

List of Experiments:

1. Preprocess textual data: Tokenization, Lemmatization, Stopword removal
2. Build a Naïve Bayes classifier for SMS/email spam detection
3. Implement a sentiment analysis model using logistic regression
4. Apply TF-IDF and Bag of Words on a corpus and analyze vector results
5. Train a Named Entity Recognition (NER) model using spaCy
6. Build a text summarizer using frequency or transformer-based models
7. Create a text classification pipeline using SVM and scikit-learn
8. Implement Word2Vec/GloVe embeddings for text similarity tasks
9. Develop a basic chatbot using NLTK / Transformers (e.g., BERT)
10. Train a Bi-LSTM for sequence labeling tasks (e.g., POS tagging)
11. Perform topic modeling using LDA (Latent Dirichlet Allocation)
12. Mini Project: Build and deploy a full-stack NLP application

Software/Tools Required:

- Python (3.7+), Jupyter Notebook/Google Colab
- Libraries: scikit-learn, pandas, NumPy, NLTK, spaCy, gensim, TensorFlow/Keras, Hugging Face Transformers
- IDE: VS Code or Jupyter Lab
- Online platforms (optional): Hugging Face Spaces, Colab, Kaggle

Textbooks & Reference Books:

1. Jurafsky & Martin, Speech and Language Processing, 3rd Edition Draft, Pearson.
2. Steven Bird, Ewan Klein, and Edward Loper, Natural Language Processing with Python**,** O'Reilly.
3. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow**,** O'Reilly, 2022.
4. Sebastian Raschka, Python Machine Learning, Packt Publishing.
5. Delip Rao and Brian McMahan, Natural Language Processing with PyTorch, O'Reilly.