**Project Domain: E-commerce [1 month Timeline]**

**Course-end Project**

**Background:** A leading online retailer, hosts numerous products with extensive consumer reviews. This project focuses on analyzing sentiments expressed in over 34,000 reviews for Amazon brand products, such as Kindle and Fire TV Stick. The dataset comprises attributes such as brand, categories, primary categories, review titles, review text and sentiment. Sentiment is categorized into three levels: "Positive," "Negative," and "Neutral." The objective is to predict the sentiment or satisfaction of a purchase based on various features and review text for unseen data.

**Project Tasks:**

**Week 1&2: Class Imbalance Problem**

1. **Exploratory Data Analysis (EDA):**
   * Understand the characteristics of positive, negative, and neutral reviews.
   * Address the class imbalance issue by examining class counts.
2. **Feature Engineering:**
   * Transform reviews into Tf-Idf scores.
3. **Classifier Selection:**
   * Implement multinomial Naive Bayes classifier.
   * Recognize challenges due to class imbalance.
4. **Tackling Class Imbalance:**
   * Apply oversampling or under-sampling techniques.
5. **Evaluation Metrics:**
   * Assess model performance using precision, recall, F1-score, and AUC-ROC curve.
   * Emphasize F1-score for evaluation.
6. **Tree-based Classifiers:**
   * Utilize Random Forest and XGBoost classifiers.
   * Leverage their capabilities in handling imbalanced classes.

**Week 3 &4: Model Selection and Advanced Techniques**

1. **Multi-class SVM and Neural Nets:**
   * Implement multi-class SVMs and neural networks.
2. **Ensemble Techniques:**
   * Explore ensemble methods like XGBoost combined with oversampled multinomial Naive Bayes.
3. **Feature Engineering:**
   * Engineer a sentiment score feature and integrate it into models for performance comparison.
4. **LSTM Implementation:**
   * Apply Long Short-Term Memory (LSTM) networks with parameter tuning, including top-word, embedding length, dropout, epochs, and layers.
5. **Comparison:**
   * Compare neural network accuracy with traditional ML algorithms.
6. **Optimization:**
   * Determine optimal settings for LSTM and GRU (Gated Recurrent Units) using techniques like Grid Search, Cross-Validation, and Random Search.
7. **Topic Modelling:**
   * Cluster similar reviews, considering different aspects like device features, aesthetics, and performance.
8. **Topic Modelling Techniques:**
   * Utilize Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) for topic modeling.
9. **Deployment:**
   * Model deployment tracking- deploy the model and monitor your model. Here you can use Streamlit or Flask or other MlOps tools which can help you to track your model’s change.

This project aims to provide insights into sentiment analysis, classification techniques, and advanced modeling approaches, ultimately enhancing understanding and decision-making in e-commerce settings.