CoderOne

Data Science and Analytics

PROJECT DETAILS

HR@CODERONE.IN

SENTIMENT ANALYSIS ON MOVIE REVIEWS

THIS PROJECT AIMS TO DEVELOP A MACHINE LEARNING MODEL FOR SENTIMENT ANALYSIS ON MOVIE REVIEWS, WITH THE GOAL OF AUTOMATICALLY CLASSIFYING REVIEWS AS POSITIVE OR NEGATIVE BASED ON THE SENTIMENT EXPRESSED IN THE TEXT. THE PROJECT INVOLVES DATA COLLECTION FROM SOURCES SUCH AS IMDB OR ROTTEN TOMATOES, PREPROCESSING THE TEXT DATA BY REMOVING NOISE AND NORMALIZING TEXT, SELECTING AND TRAINING MACHINE LEARNING MODELS SUCH AS LOGISTIC REGRESSION OR SUPPORT VECTOR MACHINES, AND EVALUATING THEIR PERFORMANCE USING METRICS LIKE ACCURACY AND F1-SCORE. THE DEPLOYED MODEL WILL PROVIDE INSIGHTS INTO AUDIENCE SENTIMENT TOWARDS MOVIES, BENEFITING FILMMAKERS, STUDIOS, AND REVIEW WEBSITES BY INFORMING DECISION-MAKING PROCESSES AND UNDERSTANDING AUDIENCE PREFERENCES.

TEXT SUMMARIZATION

THE TEXT SUMMARIZATION PROJECT AIMS TO DEVELOP A SYSTEM THAT CAN GENERATE CONCISE AND COHERENT SUMMARIES OF LONG TEXT DOCUMENTS, CAPTURING THE MAIN IDEAS AND KEY INFORMATION WHILE PRESERVING THE ORIGINAL MEANING AND CONTEXT. THE PROJECT INVOLVES PREPROCESSING THE TEXT DATA BY TOKENIZATION AND CLEANING, SELECTING APPROPRIATE SUMMARIZATION TECHNIQUES SUCH AS EXTRACTIVE OR ABSTRACTIVE METHODS, AND EVALUATING THE GENERATED SUMMARIES FOR COHERENCE AND ACCURACY. THE DEPLOYED SYSTEM WILL ENABLE USERS TO SUMMARIZE LARGE VOLUMES OF TEXT EFFICIENTLY, FACILITATING INFORMATION RETRIEVAL, DOCUMENT UNDERSTANDING, AND DECISION-MAKING ACROSS VARIOUS DOMAINS SUCH AS NEWS AGGREGATION, DOCUMENT SUMMARIZATION, AND MEETING TRANSCRIPTS ANALYSIS.

USER QUESTION ANSWERING

THE QUESTION ANSWERING SYSTEM PROJECT AIMS TO DEVELOP A MACHINE LEARNING MODEL CAPABLE OF AUTOMATICALLY ANSWERING QUESTIONS BASED ON GIVEN PASSAGES OR DOCUMENTS. THIS PROJECT INVOLVES COLLECTING A DATASET OF QUESTION-ANSWER PAIRS ALONG WITH RELEVANT PASSAGES, PREPROCESSING THE TEXT DATA TO EXTRACT FEATURES AND ENCODE CONTEXTUAL INFORMATION, SELECTING AND FINE-TUNING PRE-TRAINED MODELS SUCH AS BERT OR SIMILAR TRANSFORMER ARCHITECTURES, AND EVALUATING THE MODEL'S PERFORMANCE USING METRICS LIKE ACCURACY AND F1-SCORE. THE DEPLOYED SYSTEM WILL PROVIDE USERS WITH AN EFFICIENT AND ACCURATE MEANS OF RETRIEVING INFORMATION FROM LARGE TEXTUAL SOURCES, WITH APPLICATIONS IN FIELDS SUCH AS EDUCATION, CUSTOMER SUPPORT, AND INFORMATION RETRIEVAL.