This project focuses on building an automated speech recognition (ASR) system for digit classification using deep learning techniques. The goal is to process and interpret spoken digits (0-9) from audio recordings, transforming them into their respective numeric representations with high accuracy. The system employs Mel-Frequency Cepstral Coefficients (MFCCs) to extract features from speech signals and a Long Short-Term Memory (LSTM) neural network to model temporal dependencies within the audio data. The proposed approach was trained on a dataset of labeled audio recordings and achieved a validation accuracy exceeding 89%. The model’s performance demonstrates its capability to serve as a foundational tool for speech-based interfaces in applications such as automated customer service, voice-controlled systems, and accessibility technologies. This report details the methodology, implementation, evaluation, and potential improvements for the ASR system.