**Project Report: Masked Word Prediction Model**

**Introduction**

This project involves developing a masked word prediction model using a deep learning approach. The goal is to predict missing words in sentences based on context, which is a fundamental task in natural language processing. The model is trained on a dataset of sentences with masked words and evaluated on a separate test set.

**Dataset Description**

The dataset consists of two main files:

* **Train Data.csv**: Contains full sentences used for training the model.
* **Test Datas.csv**: Contains sentences with masked words that need to be predicted.

**Methodology**

1. **Data Preprocessing**:
   * Tokenization: Sentences were tokenized using the Tokenizer from TensorFlow.
   * Padding: Sequences were padded to a maximum length of 256 tokens.
   * Masking: Words were randomly masked in the training data to create input-output pairs.
2. **Model Architecture**:
   * The model uses a sequential architecture with the following layers:
     + **Embedding Layer**: Utilizes pre-trained GloVe embeddings with a dimension of 100.
     + **Bidirectional GRU Layers**: Two layers with 256 and 128 units, respectively, to capture contextual information.
     + **Dense Layers**: A final dense layer with softmax activation to predict the masked word.
3. **Training**:
   * The model was trained using the Adam optimizer with a learning rate of 0.001.
   * Early stopping was implemented to prevent overfitting.

**Results**

* **Training Accuracy**: The model achieved a training accuracy of approximately 0.15.
* **Validation Accuracy**: The validation accuracy plateaued at around 0.15 as well.
* **Submission File**: Predictions were made on the test set and saved to a submission.csv file.

**Evaluation Metrics**

1. **Accuracy**: This is the primary metric used to evaluate the model's performance. It measures the proportion of correctly predicted masked words out of all predictions made.

**Rationale for Accuracy**: Accuracy is crucial because it directly reflects how well the model understands the context and predicts the correct words. A higher accuracy indicates better performance.

1. **Loss**: The sparse categorical cross-entropy loss is used to train the model. Monitoring loss helps ensure that the model is learning effectively and not overfitting.

**Rationale for Loss**: Loss is important because it guides the optimization process. A decreasing loss indicates that the model is improving its predictions, while a plateauing or increasing loss may suggest overfitting or underfitting.

**Discussion**

The model's performance could be improved by adjusting hyperparameters, increasing the dataset size, or using more advanced architectures like transformers. However, the current approach provides a baseline for further enhancements.

**Conclusion**

This project demonstrates a basic approach to masked word prediction using GRU layers. While there is room for improvement, the model provides a foundation for exploring more complex NLP tasks.

**Future Work**

* **Hyperparameter Tuning**: Experiment with different learning rates, batch sizes, and optimizers.
* **Model Architecture Enhancements**: Consider using transformer models or additional GRU layers.
* **Data Augmentation**: Apply techniques like synonym replacement to increase dataset diversity.