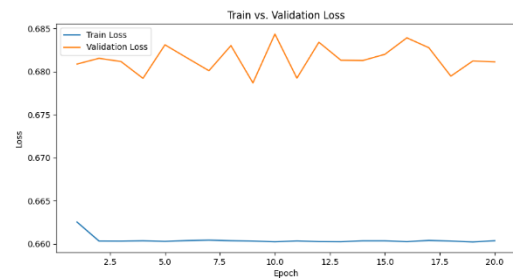
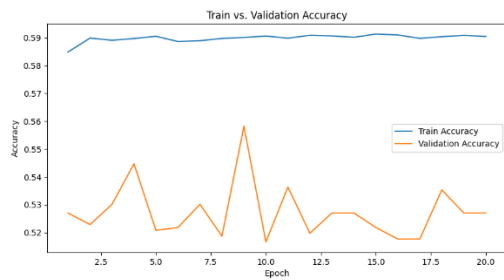


Model Results

One-Hot log-linear Model



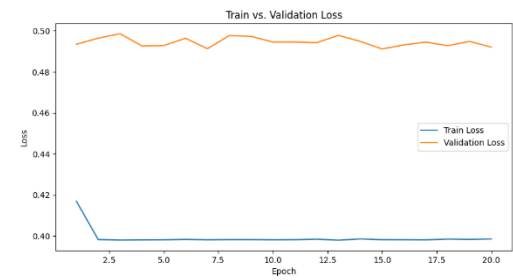
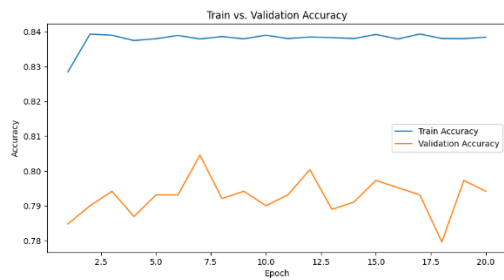
Test Loss: 0.6783

Test Accuracy: 0.5291

Negated Polarity - Loss: 0.6964, Accuracy: 0.4839

Rare Words - Loss: 0.7321, Accuracy: 0.3000

Word2Vec log-linear Model



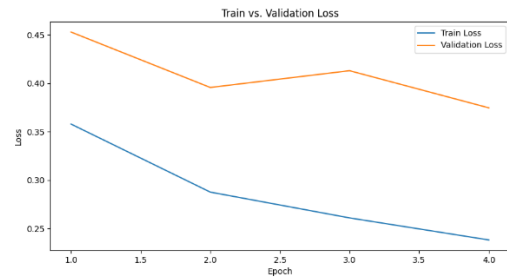
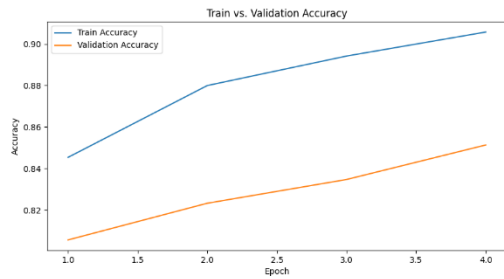
Test Loss: 0.4683

Test Accuracy: 0.8191

Negated Polarity - Loss: 0.7546, Accuracy: 0.5645

Rare Words - Loss: 0.6506, Accuracy: 0.7400

Word2Vec LSTM Model



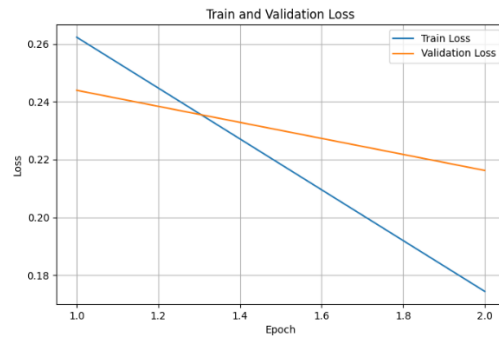
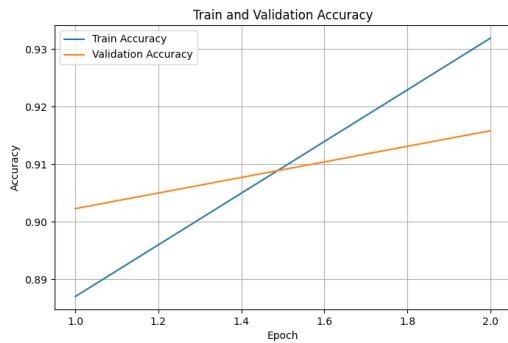
Test Loss: 0.3621

Test Accuracy: 0.8482

Negated Polarity - Loss: 0.9091, Accuracy: 0.6613

Rare Words - Loss: 0.6503, Accuracy: 0.7800

Transformer Model



Test Loss: 0.1905

Test Accuracy: 0.9085

Negated Polarity - Accuracy: 0.8226

Rare Words - Accuracy: 0.8600

Comparing the different models

1. The Word2Vec log-linear model significantly outperforms the simple log-linear model this might be because the simple log-linear model uses one-hot embeddings, which fail to capture semantic relationships between words. Where the Word2Vec embeddings encode richer semantic information, enabling the model to generalize better.
2. LSTMs model sequential dependencies, capturing context across words in a sentence. This capability helps LSTMs perform better on more complex patterns, such as negated polarity sentences. Transformers leverage attention mechanisms to capture long-range dependencies and relationships between all words in a sentence, regardless of distance. This explains the superior performance of Transformers on both test and subset accuracies, especially for challenging cases like negated polarity and rare words.
3. Word2Vec embeddings provide a substantial improvement over one-hot embeddings due to richer semantic information. Sequential models like LSTM further enhance performance by capturing dependencies across tokens. Transformers outperform all other models due to their ability to model complex relationships and capture global context, excelling particularly on challenging subsets like negated polarity and rare words.
 - **Negated Polarity:** This subset is challenging because it requires understanding nuanced relationships, such as the impact of negation. Simple models (e.g., one-hot embeddings) cannot capture this complexity, while Transformers excel due to their global attention mechanism.
 - **Rare Words:** The simple log-linear model struggles as one-hot embeddings cannot generalize beyond specific word indices. Models with pre-trained embeddings (Word2Vec, LSTM, Transformer) perform better, with Transformers achieving the highest accuracy by capturing contextual meanings effectively.