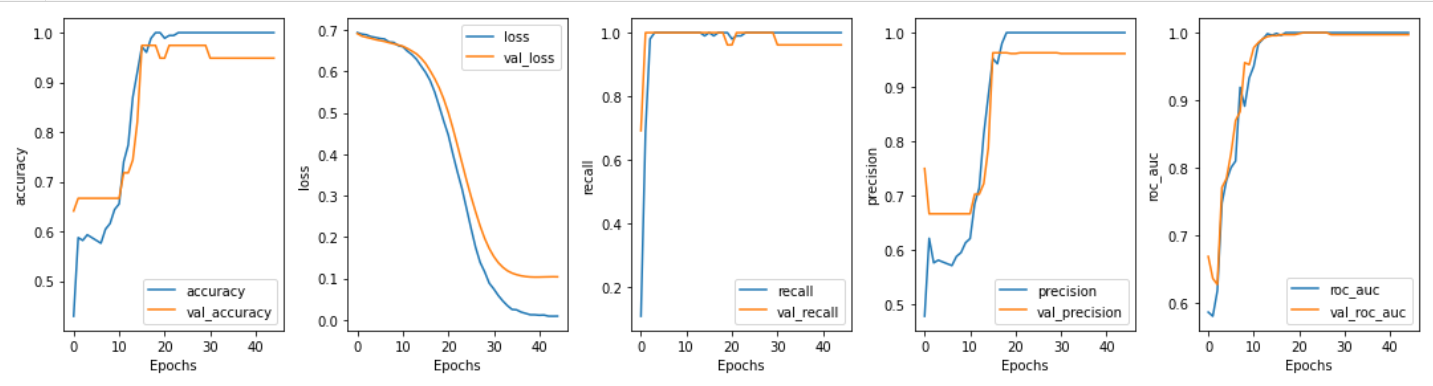
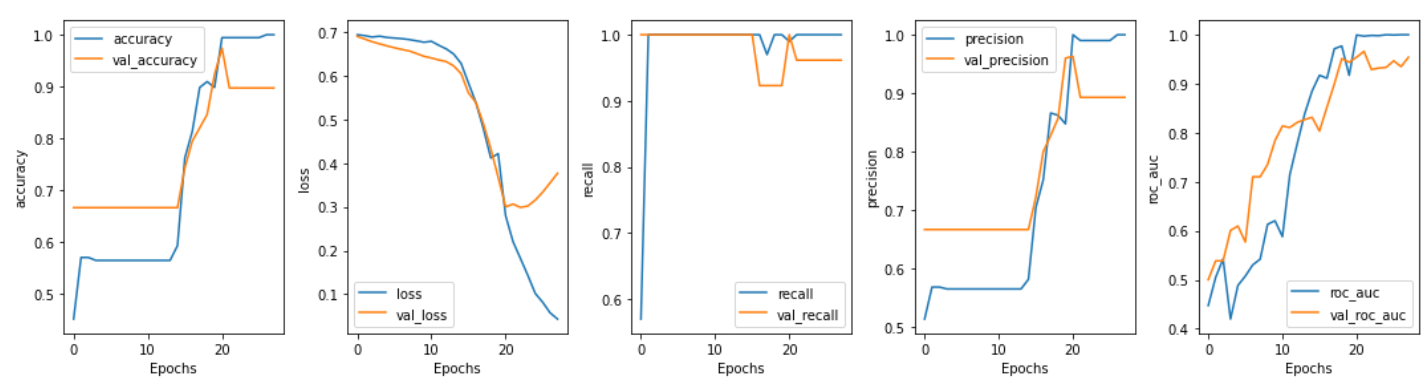
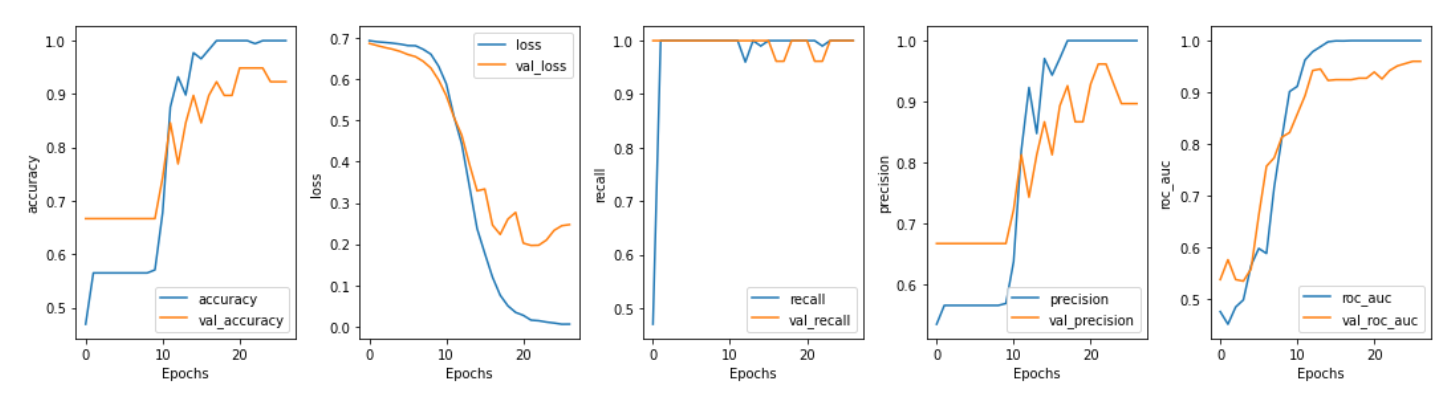
**Positive and Negative Reviews Classification**

**Data Preprocessing:** I started by merging the review texts and labels into a single dataset. I then preprocessed the text data by converting it to lowercase, tokenizing it, removing punctuation and common English stopwords and applying stemming to the remaining tokens. The sentiment labels were also converted into binary values, where 'Positive' became 1 and 'Negative' became 0.

**Model Selection and Training:** I explored three different models for text classification: Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (LSTM) network and an LSTM-CNN hybrid. All models were trained with a vocabulary size of 10,000 words and a maximum sequence length of 120. I utilized callback functions such as model checkpointing, early stopping, and learning rate reduction to prevent overfitting. I decided to simplify the models to reduce training time.

**Model Performance:**

* The CNN model achieved good results with a focus on balanced accuracy: 
* The LSTM model, although promising, was prone to overfitting: 
* The LSTM-CNN hybrid model:



**Challenges:** Due to the limited size of our dataset, overfitting was a significant challenge. To mitigate this, I selected hyperparameters, used regularization techniques such as dropout and closely monitored validation loss during training.

**Conclusion:** In conclusion, building an efficient and accurate text classification model on a small dataset proved to be a challenging yet rewarding task. Chosen CNN model, despite the challenges, provided a balance between performance and generalization, demonstrating that even with limited resources, creative model architectures and thoughtful training strategies can yield promising results.

Начало формы