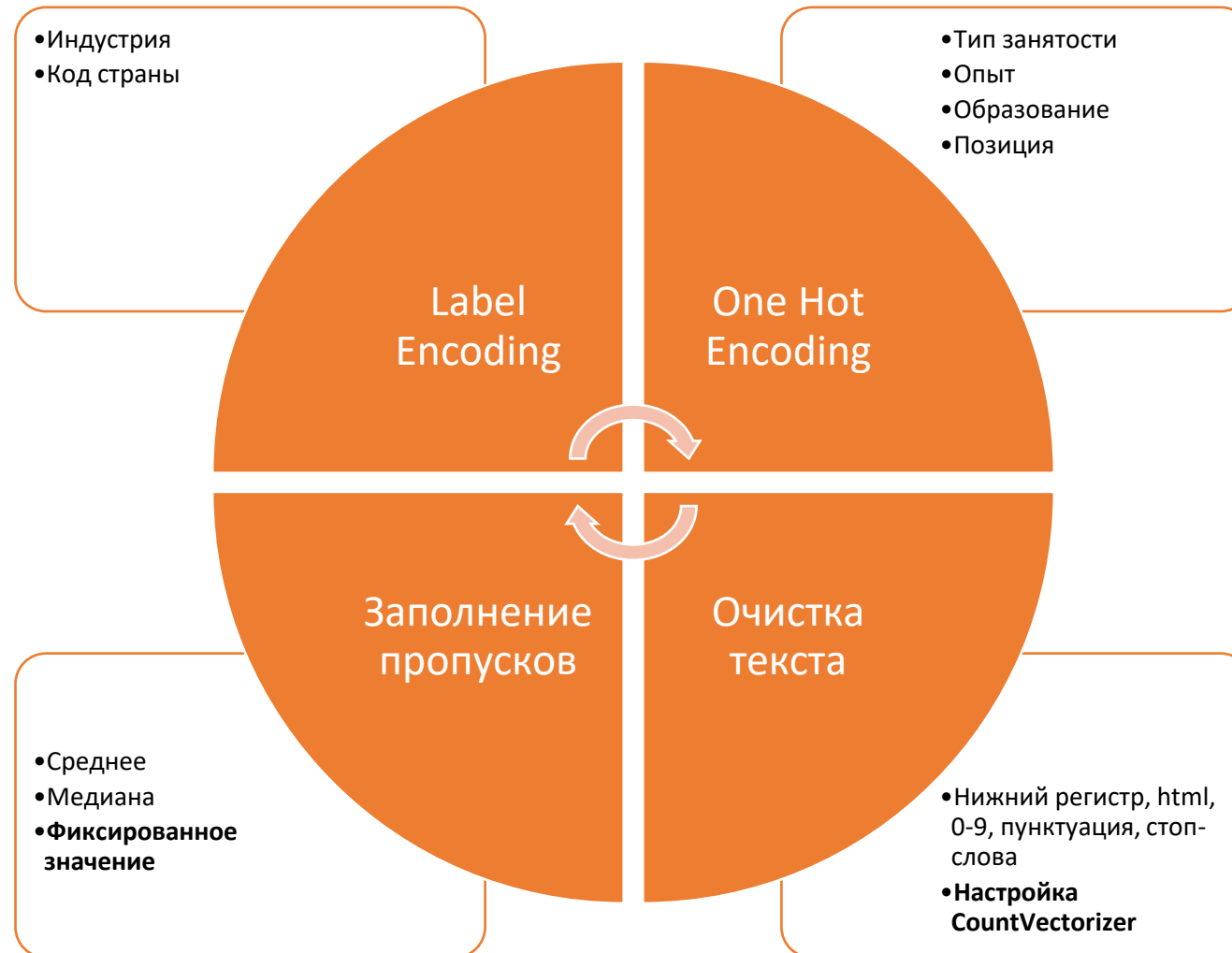


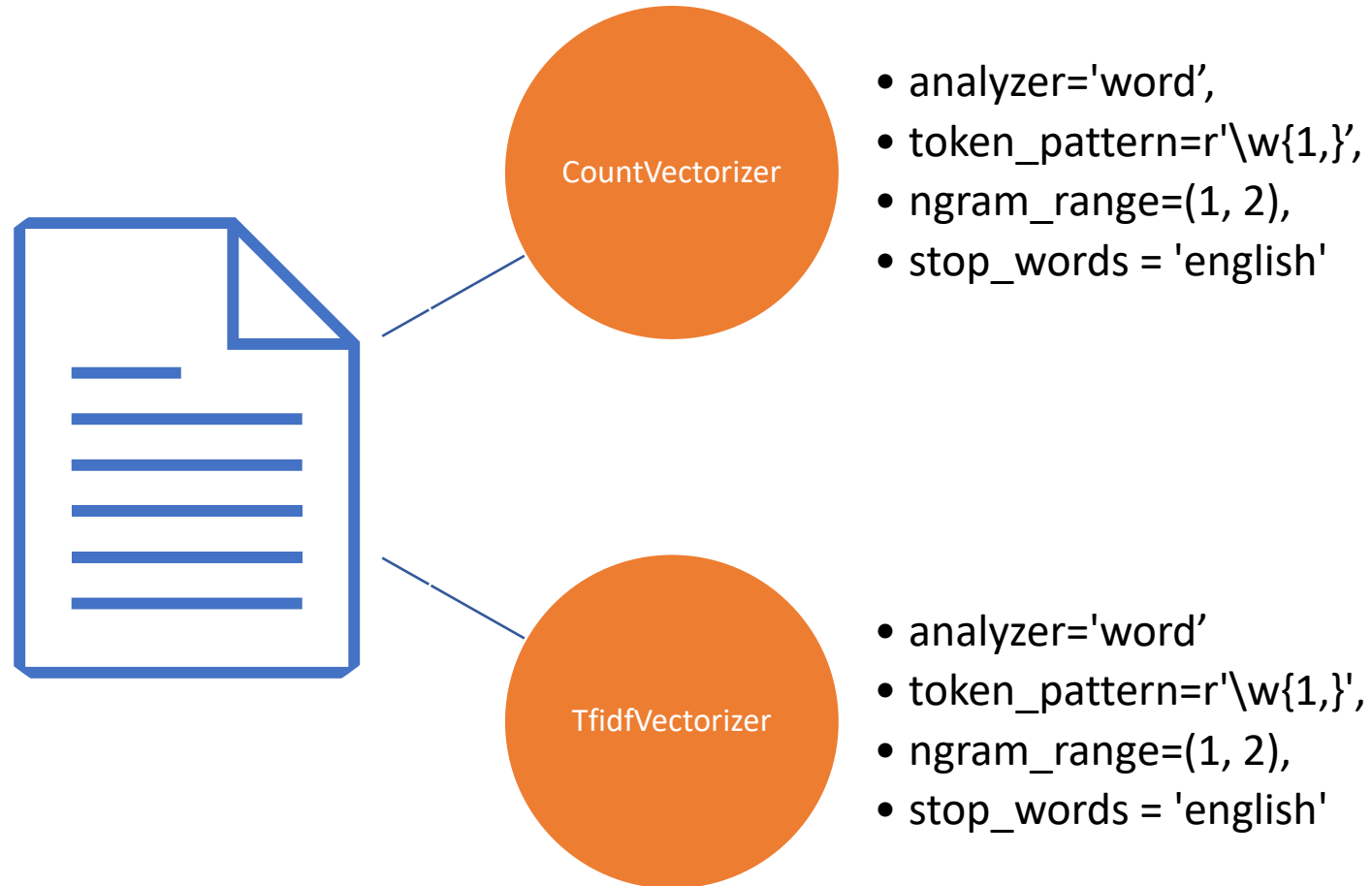
RuCode Fake Job Postings



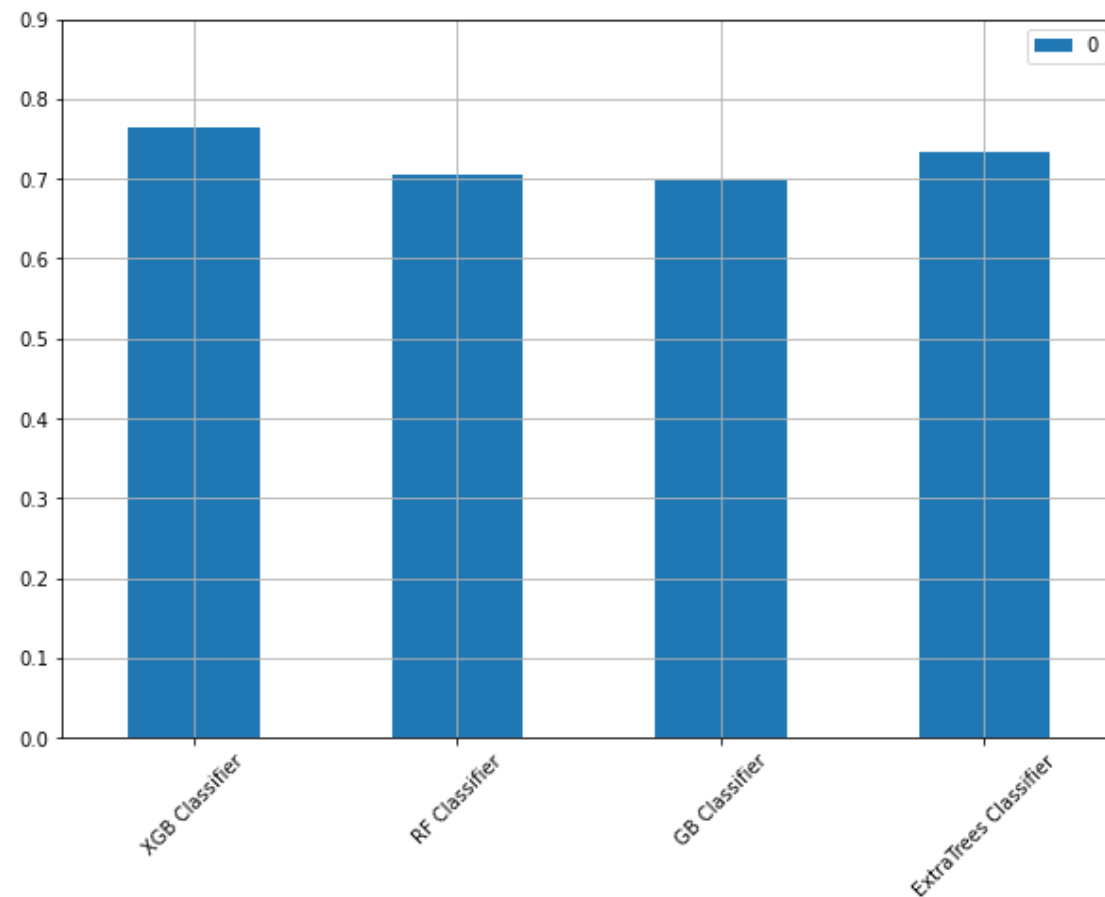
Предобработка



Предобработка текстовых признаков



Базовые алгоритмы

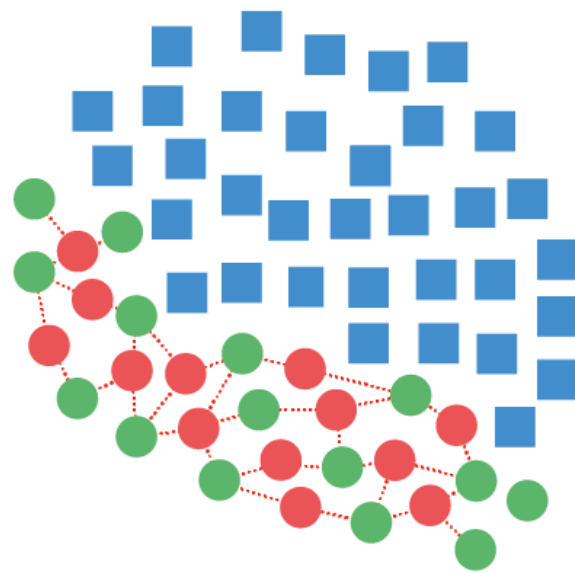


Oversampling

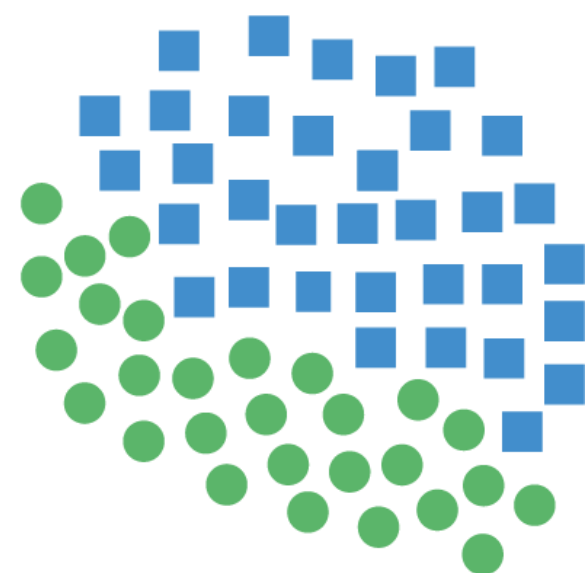
Synthetic Minority Oversampling Technique



Original Dataset



Generating Samples



Resampled Dataset

XGBClassifier

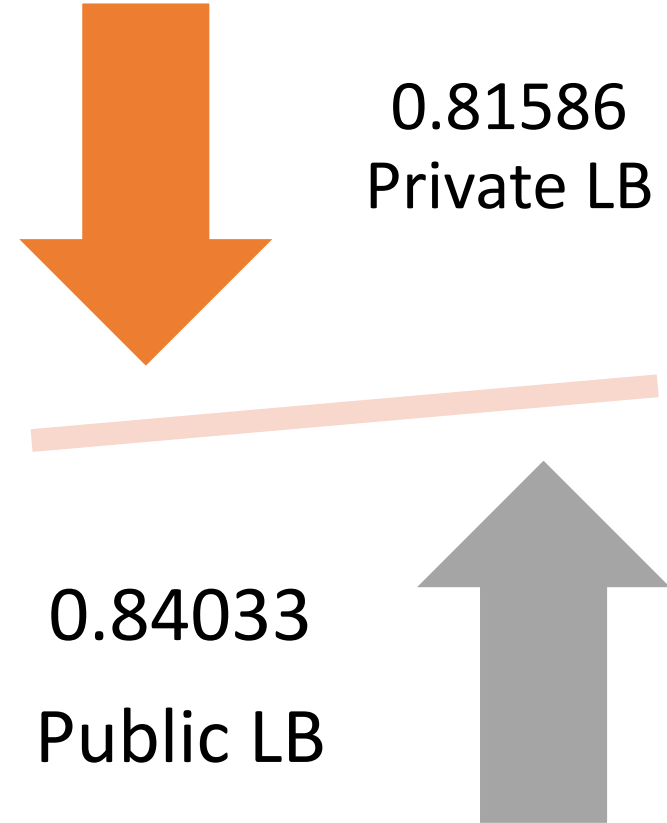
max_depth=7,

n_estimators=300,

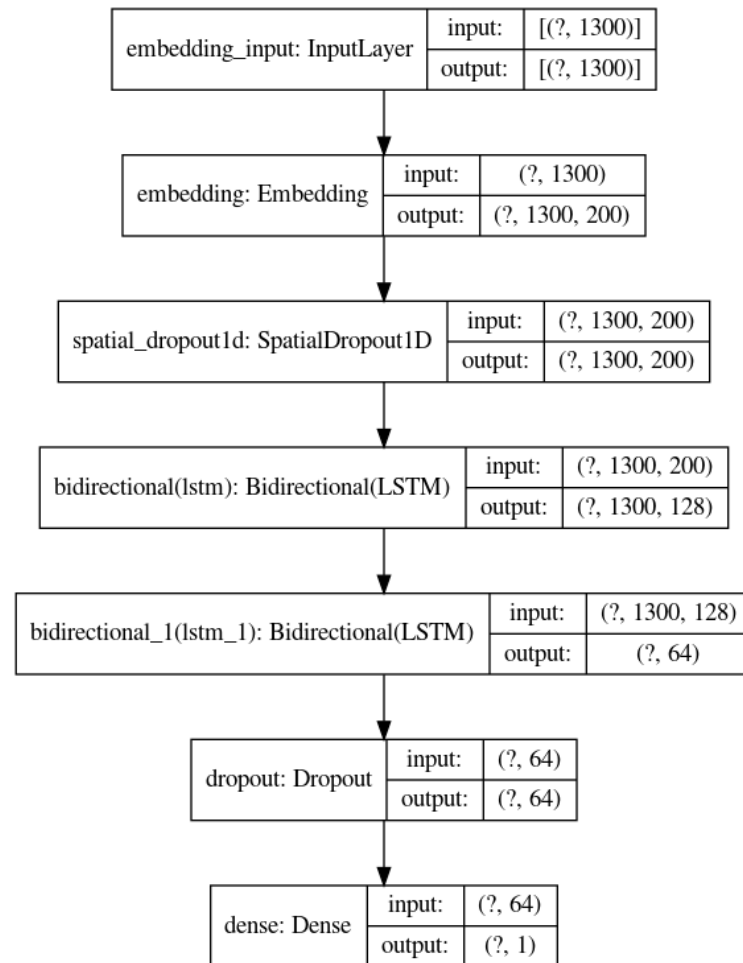
colsample_bytree=0.8,

subsample=0.8,

learning_rate=0.1



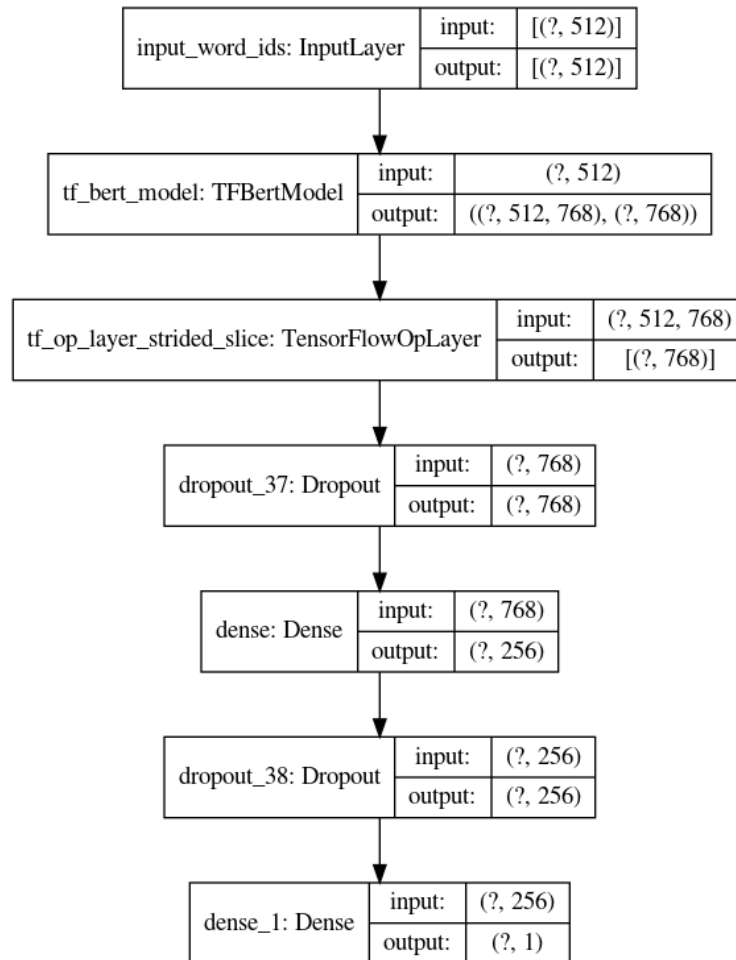
Нейросети – Bidirectional LSTM



Нейросети - BERT

- BertWordPieceTokenizer + distilbert_base_uncased/vocab.txt
- DistilBERT меньше и быстрее BERT
- Callback – F1 Score
- Metrics – Recall, Precision
- Loss – binary crossentropy
- LearningRateScheduler - Адаптация скорости обучения
- Dropout (0.35) – Dense – Dropout (0.1) – Dense

Нейросети - BERT

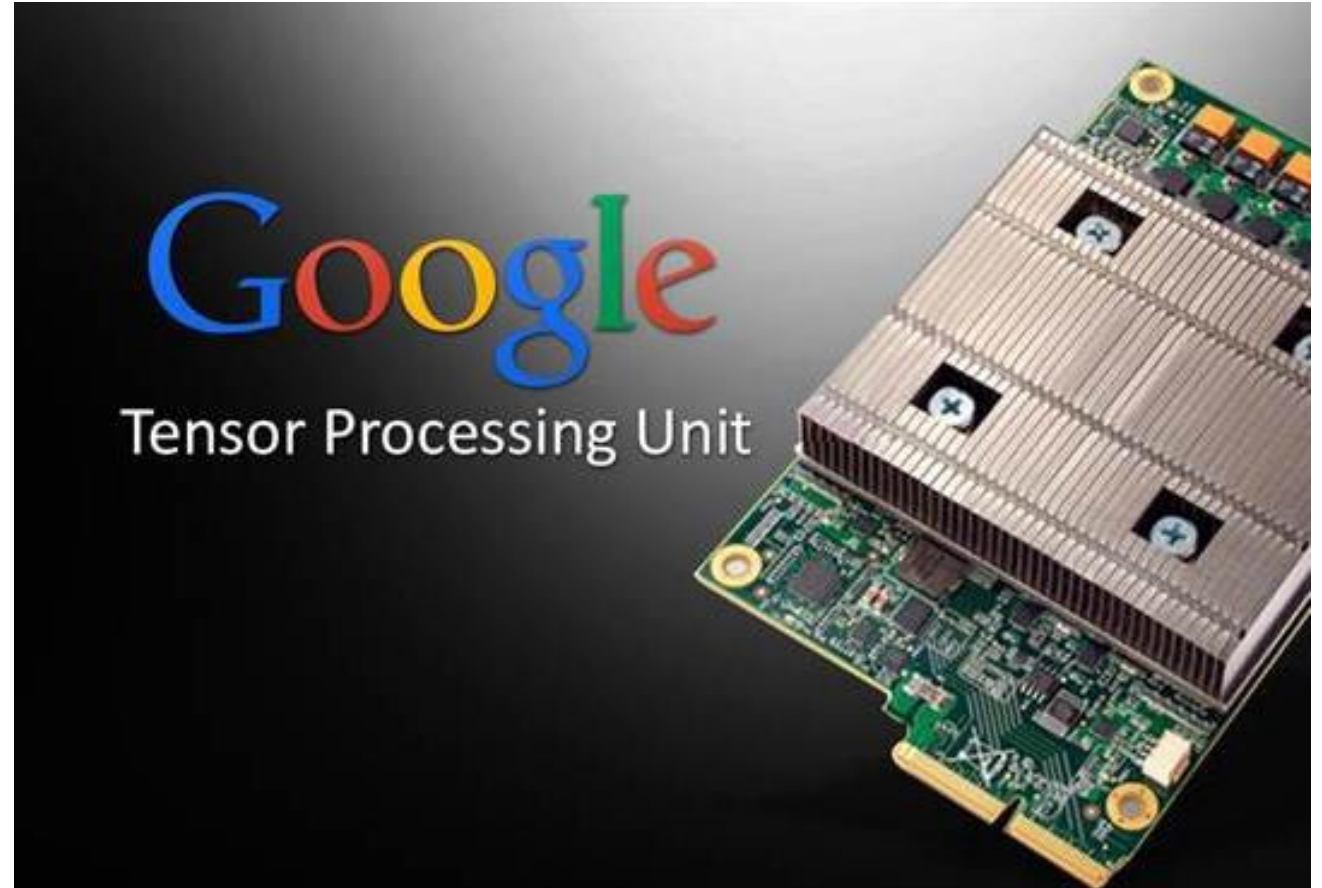


Результаты



TPU - Тензорный процессор Google

- По сравнению с графическими процессорами, рассчитан на более высокий объём вычислений с пониженной точностью при более высоких производительности на ватт и отсутствии модуля для растризации и текстурных блоков.



Спасибо за внимание!

