Algorithm 1 Feature Extraction

Define Extract Features Function extract_featurestext
words = text.split()
num_words = len(words)
num_sentences = len(nltk.sent_tokenize(text))
avg_word_length = sum(len(word) for word in words) / num_words if num_words ¿ 0 else 0
unique_words = set(words)
vocabulary_richness = len(unique_words) / num_words if num_words ¿ 0 else 0
pos_tags = nltk.pos_tag(words)
noun_count = sum(1 for tag in pos_tags if tag[1].startswith('N'))
verb_count = sum(1 for tag in pos_tags if tag[1].startswith('V'))
adjective_count = sum(1 for tag in pos_tags if tag[1].startswith('J'))
return { 'num_words': num_words, 'num_sentences': num_sentences, 'avg_word_length': avg_word_length, 'vocabulary_richness': vocabulary_richness, 'noun_count': noun_count, 'verb_count': verb_count, 'adjective_count': adjective_count }
Extract Features
features = [extract_features(text) for text in preprocessed_data]