AUTOMATED ESSAY SCORING

The quality of an essay is affected by the following four primary dimensions:

* Topic relevance
* Grammar
* Mechanics – including spelling rules, punctuations, capitalization rules, etc.
* Word usage and writing style
* Sentence complexity
* Organization and coherence – essay structure and appropriate transition
* Ideas and examples
* Tone and persuasiveness
* Thesis clarity

How to deal with these:

1. **Topic relevance:**

- Use topic modelling techniques to ensure the essay addresses the main themes or keywords of the given topic.

- Implement algorithms to measure semantic similarity between the essay content and the expected topic.

1. Topic Modelling:

- Utilize techniques like Latent Dirichlet Allocation (LDA) or Non-Negative Matrix Factorization (NMF) to identify the main themes or topics within the essay.

- Compare the distribution of topics in the essay with the expected topic or topics provided in the prompt.

- Assign a relevance score based on the degree of overlap between the topics identified in the essay and the expected topic.

2. Semantic Similarity:

- Calculate the semantic similarity between the essay content and the keywords or concepts provided in the prompt.

- Utilize pre-trained word embeddings such as Word2Vec, GloVe, or BERT to represent the essay and prompt text in a semantic vector space.

- Measure the cosine similarity or other distance metrics between the essay vector and the prompt vector to gauge relevance.

3. Keyword Matching:

- Identify key terms or keywords in the prompt that define the main focus or requirements of the topic.

- Search for occurrences of these keywords within the essay and assess their frequency and context.

- Assign a relevance score based on the presence and usage of relevant keywords in the essay.

4. Supervised Learning:

- Train a machine learning classifier on annotated data to predict the relevance of essays to specific topics.

- Use features such as essay content, prompt text, and keyword occurrences as input to the classifier.

- Incorporate feedback from human raters to refine the model's predictions and improve its accuracy.

5. Topic-Specific Language Models:

- Fine-tune pre-trained language models such as GPT or BERT on topic-specific corpora.

- Use the fine-tuned models to generate embeddings or assess the relevance of essays to the given topic.

- Fine-tuning allows the model to capture domain-specific nuances and improve its ability to evaluate topic relevance.

6. Ensemble Approaches:

- Combine multiple methods mentioned above into an ensemble model to leverage their complementary strengths.

- Weight the predictions of individual models based on their performance on a validation set or through meta-learning techniques.

- Ensemble models often provide more robust and accurate assessments of topic relevance by mitigating the weaknesses of individual approaches.

2. **Grammar:**

- Employ grammatical parsing techniques to detect and correct errors in sentence structure.

- Train a model on annotated data to identify grammatical errors such as subject-verb agreement, tense consistency, and pronoun reference.

3. **Mechanics:**

- Develop algorithms to check spelling, punctuation, and capitalization errors.

- Use rule-based systems or machine learning models trained on annotated data to identify and correct mechanical mistakes.

4. **Word usage and writing style:**

- Utilize word embeddings or pre-trained language models to assess the appropriateness and sophistication of vocabulary.

- Train a model to evaluate writing style based on features like sentence length, use of passive voice, and variation in sentence structure.

5. **Sentence complexity:**

- Apply syntactic analysis to measure the complexity of sentence structures.

- Use readability metrics such as Flesch-Kincaid Grade Level or Gunning Fog Index to gauge the complexity of the essay.

6. **Organization and coherence:**

- Develop algorithms to analyze the logical flow of ideas and identify paragraph transitions.

- Train a model to recognize cohesive devices such as pronouns, transitional phrases, and conjunctions.

7. **Ideas and examples:**

- Use topic modeling and sentiment analysis to evaluate the depth and relevance of ideas presented in the essay.

- Train a model to identify and assess the quality of examples provided to support arguments or claims.

8. **Tone and persuasiveness:**

- Implement sentiment analysis techniques to determine the overall tone of the essay.

- Train a model to recognize persuasive language and rhetorical devices commonly used in persuasive writing.

9. Thesis clarity:

- Develop algorithms to extract the main thesis or argument from the essay.

- Train a model to assess the clarity and specificity of the thesis statement.

For each criterion, you'll need to collect and annotate a large dataset of essays with human scores for training and evaluation purposes. Additionally, fine-tuning the models and continuously updating them with new data will help improve the accuracy and effectiveness of your AES system over time.