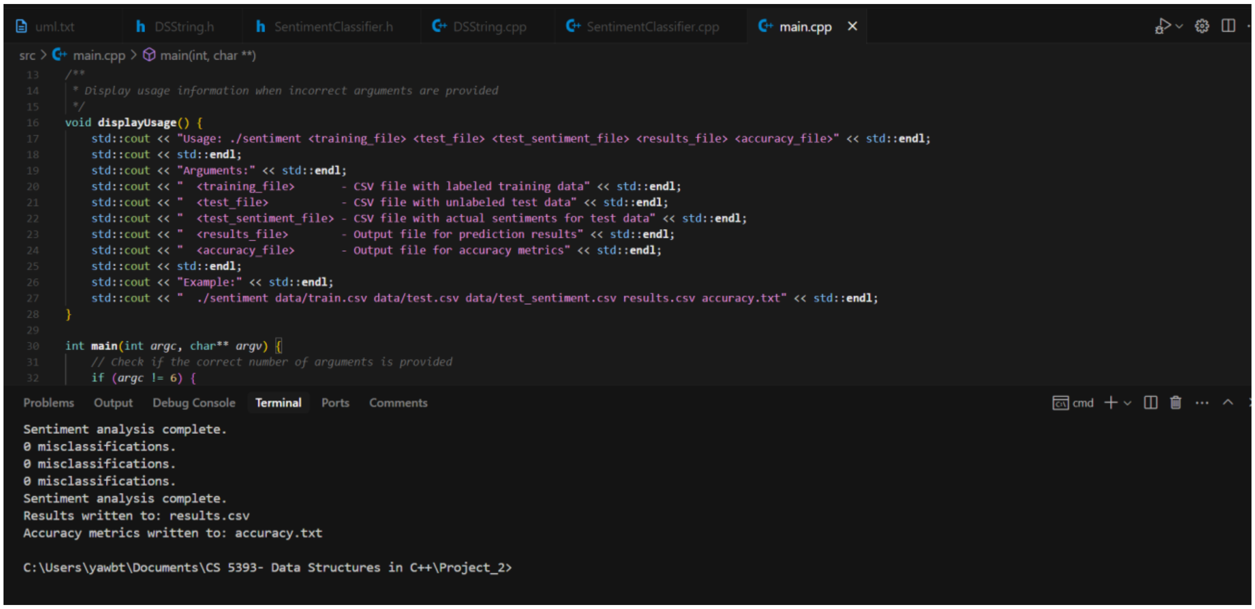
Yaw Boateng, CS 5393 Section 002

Project 2



```cpp
/**
 * Display usage information when incorrect arguments are provided
 */
void displayUsage() {
    std::cout << "Usage: ./sentiment <training_file> <test_file> <test_sentiment_file> <results_file> <accuracy_file>" << std::endl;
    std::cout << std::endl;
    std::cout << "Arguments:" << std::endl;
    std::cout << "  <training_file>       - CSV file with labeled training data" << std::endl;
    std::cout << "  <test_file>           - CSV file with unlabeled test data" << std::endl;
    std::cout << "  <test_sentiment_file> - CSV file with actual sentiments for test data" << std::endl;
    std::cout << "  <results_file>        - Output file for prediction results" << std::endl;
    std::cout << "  <accuracy_file>       - Output file for accuracy metrics" << std::endl;
    std::cout << std::endl;
    std::cout << "Example:" << std::endl;
    std::cout << "  ./sentiment data/train.csv data/test.csv data/test_sentiment.csv results.csv accuracy.txt" << std::endl;
}

int main(int argc, char** argv) {
    // Check if the correct number of arguments is provided
    if (argc != 6) {
```

```
Sentiment analysis complete.
0 misclassifications.
0 misclassifications.
0 misclassifications.
Sentiment analysis complete.
Results written to: results.csv
Accuracy metrics written to: accuracy.txt

C:\Users\yawbt\Documents\CS 5393- Data Structures in C++\Project_2>
```

This sentiment analysis project is a C++ implementation that classifies tweets as either positive or negative sentiment using a custom-built lexicon-based machine learning approach. At its core, it leverages a bespoke DSString class for memory-efficient text processing, combined with a SentimentClassifier that learns word-sentiment associations during training by analyzing frequency patterns in labeled data. The system processes tweets by tokenizing text into individual words, calculating sentiment scores based on learned word frequencies, and making binary predictions (positive/negative) through a simple yet effective statistical model. After generating predictions, the program evaluates its accuracy against ground truth data, demonstrating fundamental machine learning concepts including training, prediction, and evaluation phases without relying on external NLP libraries.