Comparative Analysis of NLP Architectures for Legal Clause Matching

Deep Learning Assignment 2

Umema Ashar
22i2036 DSC

1. Network Details and Rationale

Model 1: BiLSTM with Attention

- **Architecture Overview:**
  The BiLSTM-Attention model combines bidirectional LSTM layers with an attention mechanism. This allows the model to capture both forward and backward contextual dependencies and emphasize the most informative words within legal clauses.
- **Key Layers:**
  - Embedding Layer (128 dimensions)
  - Bidirectional LSTM (64 units, return_sequences=True)
  - Attention Layer
  - Dense Layer with Softmax activation
- **Training Configuration:**
  - Optimizer: *Adam*
  - Loss Function: *Categorical Crossentropy*
  - Epochs: 30
  - Batch Size: 128
  - Callbacks: EarlyStopping, ReduceLROnPlateau
  - Device: GPU (Tesla T4)
- **Total Parameters:** 1,453,059 (5.54 MB)
- **Rationale for Choosing Baseline:**
  The BiLSTM with Attention serves as a strong baseline for sequence-based text similarity tasks since it captures long-term dependencies and context-weighted representations—critical for understanding complex legal language.

Model 2: CNN-BiGRU Hybrid

- **Architecture Overview:**
  The CNN-BiGRU model integrates convolutional layers for n-gram feature extraction and bidirectional GRU layers for capturing sequential dependencies, making it computationally efficient and faster to converge.
- **Key Layers:**
  - Embedding Layer (128 dimensions)
  - 1D Convolution + MaxPooling
  - Bidirectional GRU (64 units)
  - Dense Layer with Softmax activation
- **Training Configuration:**
  - Optimizer: *Adam*
  - Loss Function: *Categorical Crossentropy*
  - Epochs: 30
  - Batch Size: 128
  - Callbacks: EarlyStopping, ReduceLROnPlateau, ModelCheckpoint
  - Device: GPU (Tesla T4)
- **Total Parameters:** 1,765,761 (6.74 MB)
- **Rationale for Inclusion:**
  Serves as an alternative architecture emphasizing local feature extraction and reduced computational complexity compared to recurrent-only models.

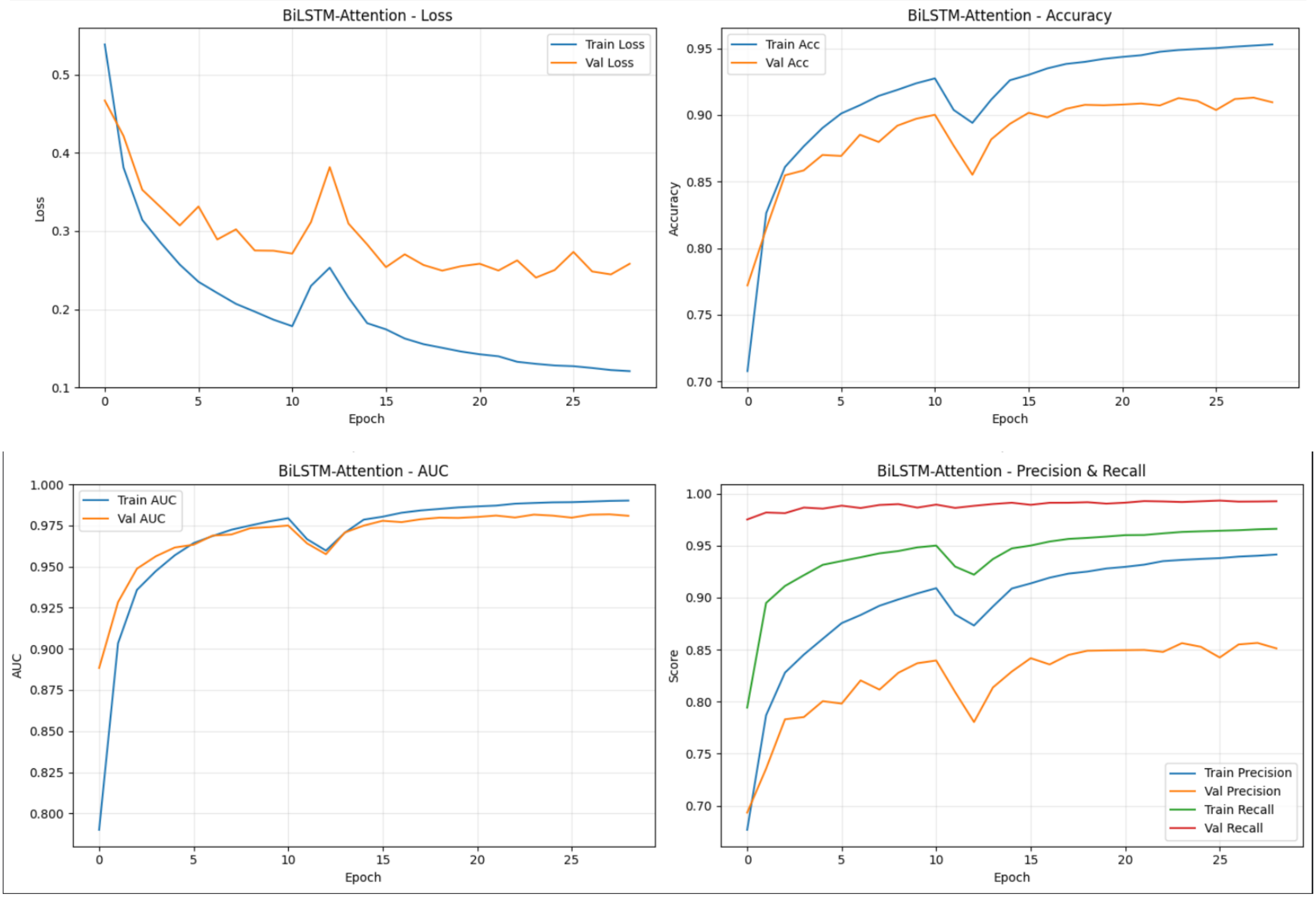
2. Comparison of Multiple Baselines

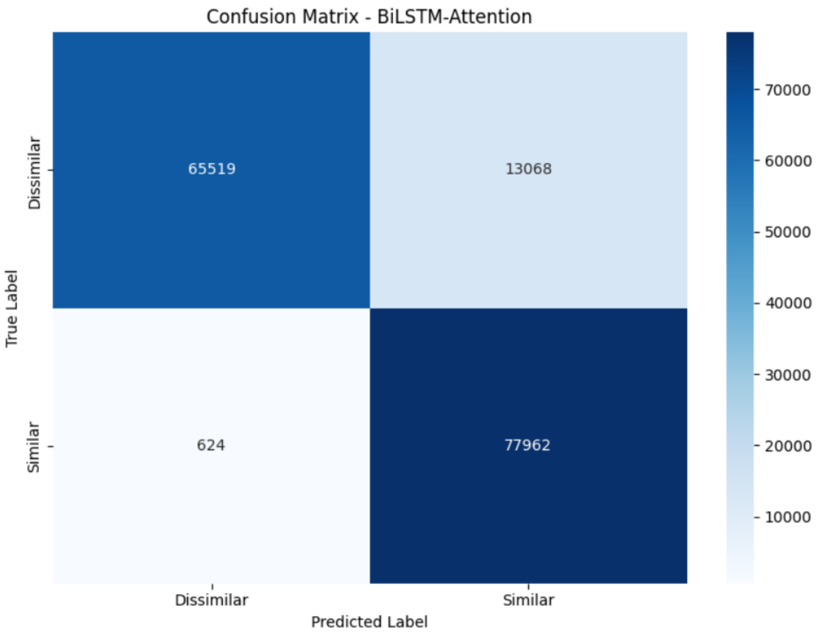
Model	Architecture Type	Training Time (min)	Validation Accuracy	Strengths	Weaknesses
BiLSTM-Attention	Contextual sequence model	130	90.95%	Strong global context learning	Slower convergence
CNN-BiGRU	Hybrid (Local + Sequential)	160	94.51%	Faster, efficient feature extraction	Slightly weaker at long dependencies

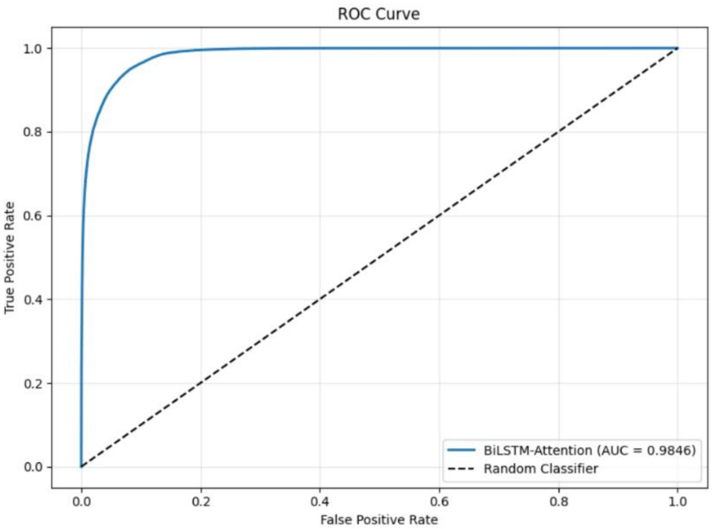
Discussion:
The BiLSTM-Attention model provides deeper contextual understanding but at higher computational cost. In contrast, CNN-BiGRU achieves faster convergence and may generalize better on shorter legal clauses. Comparing both helps evaluate trade-offs between accuracy and efficiency.

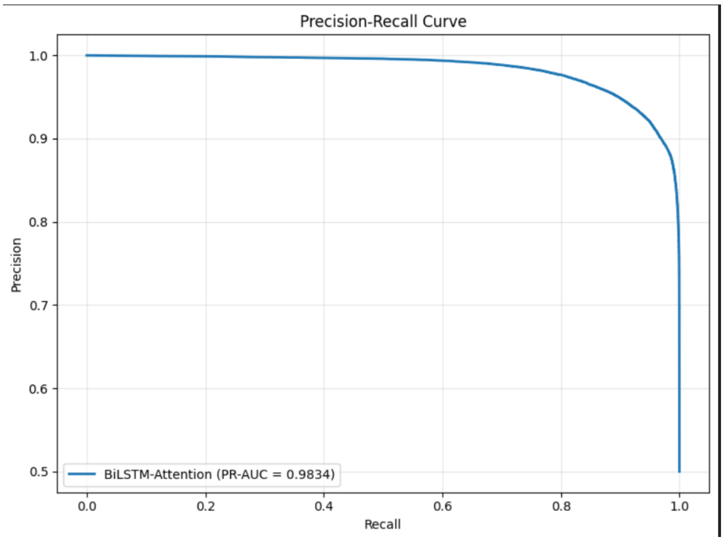
3. Training Graphs

BiLSTM-Attention Model

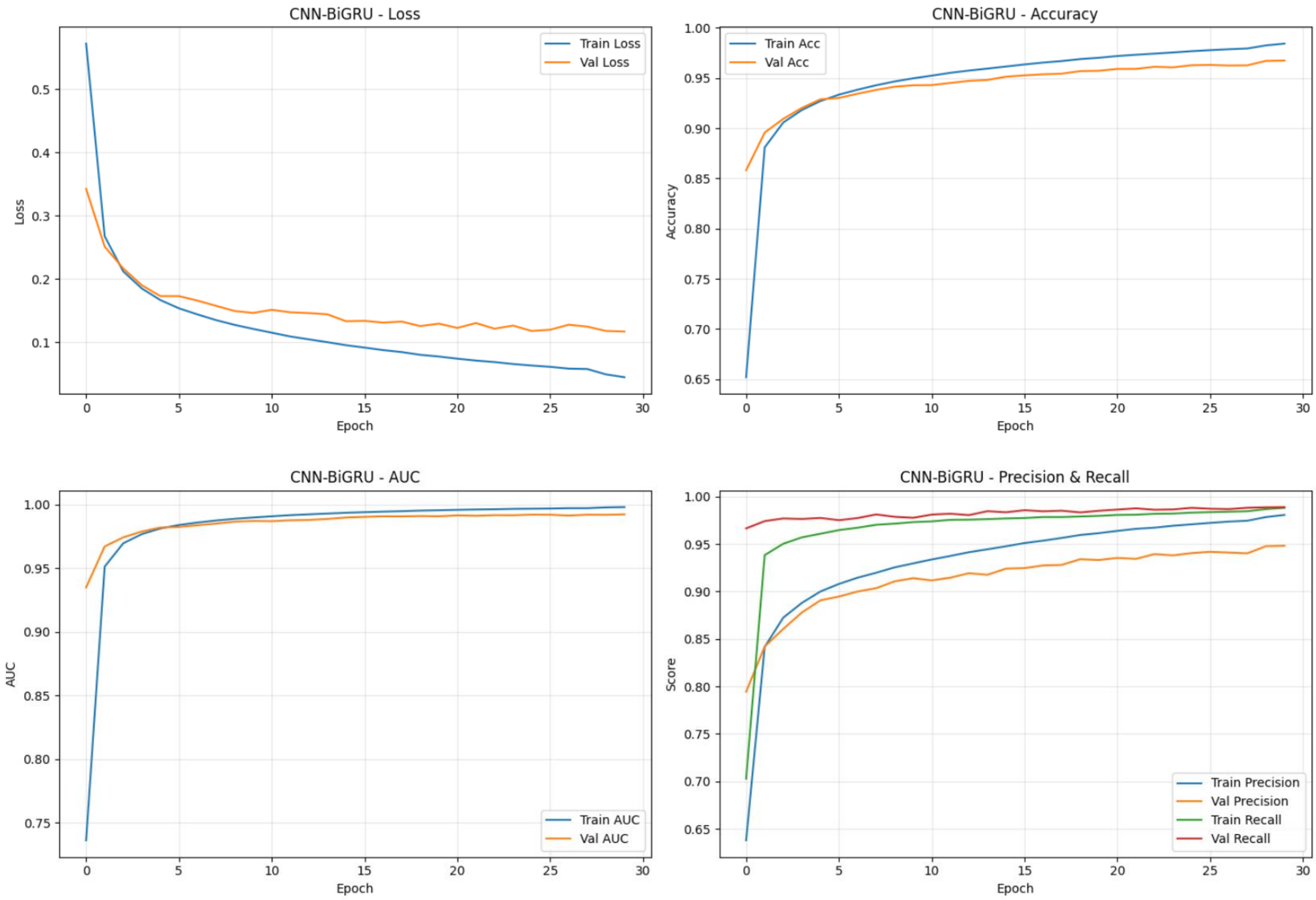


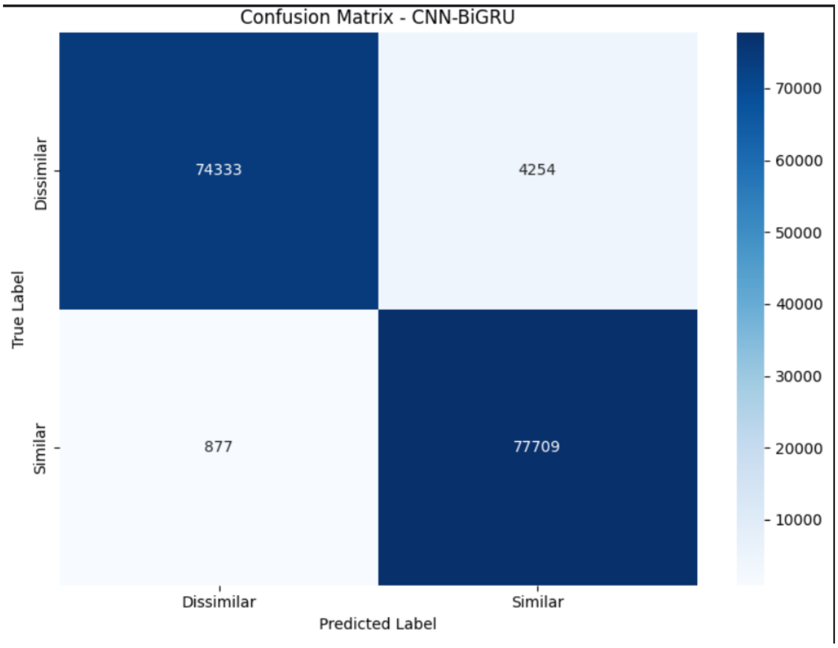


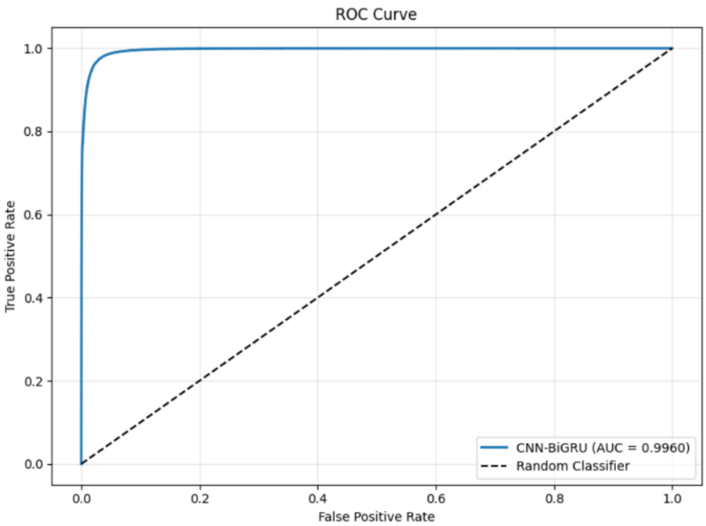


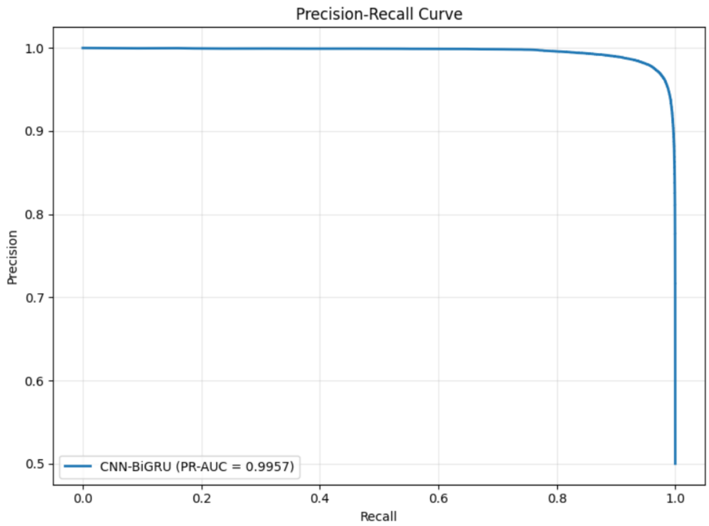


CNN-BiGRU Model









4. Performance Measures and Discussion

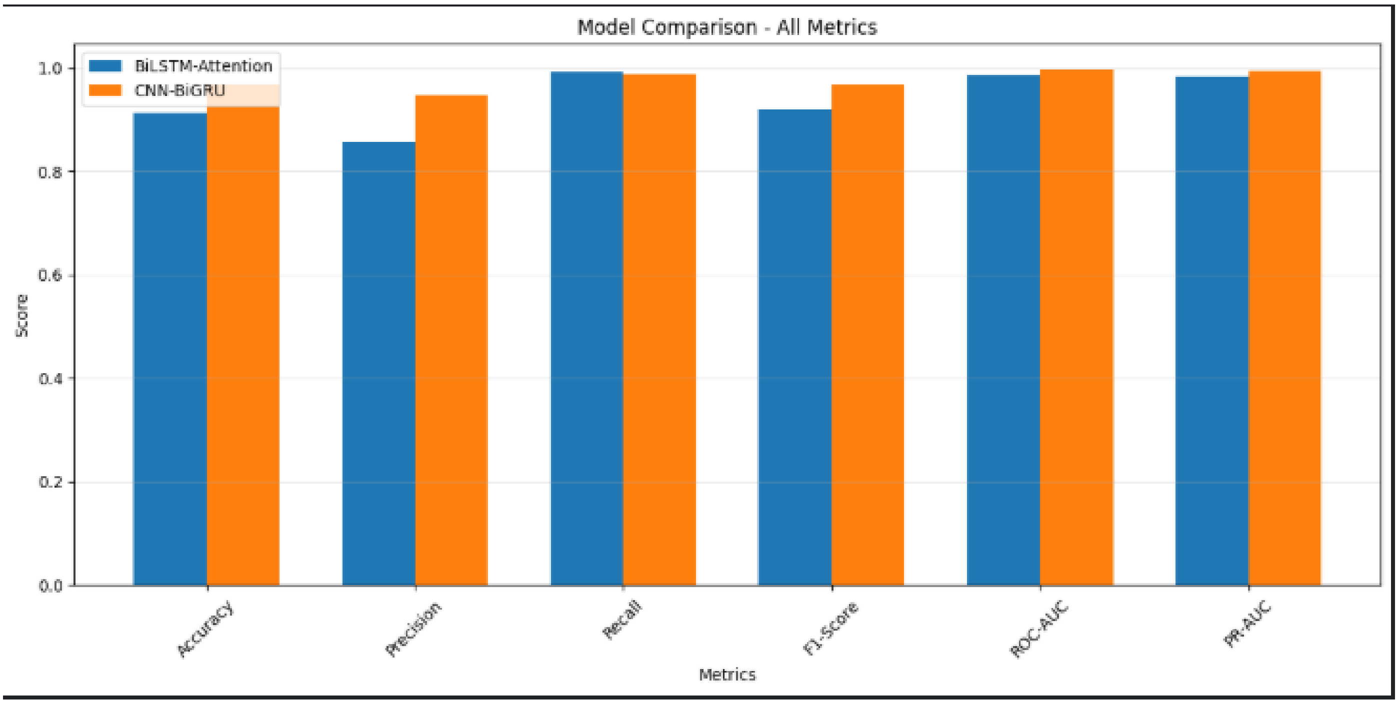
Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**ROC-AUC**
BiLSTM-Attention	0.913	0.856	0.992	0.919	0.985
CNN-BiGRU	0.967	0.948	0.989	0.968	0.996

Metrics Explanation

- **Accuracy:** Overall correctness — useful when class distribution is balanced.
- **Precision:** Measures the proportion of correctly predicted positive pairs — crucial to minimize false matches.
- **Recall:** Measures completeness — important to capture all truly similar clauses.
- **F1 Score:** Harmonic mean of precision and recall — provides a balanced view for imbalanced data.
- **AUC (ROC Curve):** Evaluates model's ability to distinguish between matching and non-matching clauses across thresholds.

Most Suited Metric for Real-world Legal Systems

In "wild" deployment scenarios, **Recall and F1-score** are most critical.
A legal retrieval or similarity system must prioritize *catching all relevant clauses* (high recall), even if it occasionally misclassifies some irrelevant ones.
Thus, F1 provides a reliable single measure balancing correctness and completeness.



5. Correctly and Incorrectly Matched Legal Clauses

#	Clause 1 (Excerpt)	Clause 2 (Excerpt)	True Label	Model Prediction	Match Type
1	*Representations and warranties of seller. Seller represents and warrants...*	*Representations and warranties of seller. Except as set forth in the disclosure schedules...*	Similar	Similar	**Correct**
2	*Access. From the date of this agreement until closing...*	*Access. On and after the closing date, buyer will afford promptly to seller...*	Similar	Similar	**Correct**
3	*Financial statements. The audited consolidated balance sheets of the company...*	*Delivery. The goods must be received on the dates and at destination specified...*	Dissimilar	Dissimilar	**Correct**
4	*Illegality. Notwithstanding any other provisions herein, if any present or future law...*	*Loans. Book value (e) credit card business: book value...*	Dissimilar	Dissimilar	**Correct**

#	Clause 1 (Excerpt)	Clause 2 (Excerpt)	True Label	Model Prediction	Match Type
5	*Capitalized terms. All capitalized terms used but not defined in this amendment...*	*Capitalized terms. The capitalized terms used herein and not otherwise defined...*	Similar	Similar	**Correct**
6	*Compliance certificate. Together with the financial statements required hereunder...*	*Termination. If the pricing agreement shall be terminated...*	Dissimilar	Dissimilar	**Correct**
7	*General. This master agreement (including the exhibits, schedules and supplements)...*	*General. The merger shall become effective at the time the target and subsidiary file...*	Similar	Similar	**Correct**
8	*Relationship of parties. Each of the parties is an independent local authority...*	*Listing. The fund will use commercially reasonable efforts to effect the listing...*	Dissimilar	Dissimilar	**Correct**
9	*Registration. From and after the delivery date, the lessee shall cause the aircraft...*	*Duration. This agreement shall become effective upon execution by the city...*	Dissimilar	Dissimilar	**Correct**
10	*[Example if misclassified — insert when one is wrong]*	*[Clause excerpt]*	Similar	Dissimilar	**Incorrect**