\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Problem Statement**: Extract MetaData from Rental Agreements

**Approach MindMap** :

**(Dividing The Problem Statement into 4 Stages)**

**STAGE - 1:**

**OBJECTIVE : SEMI-SUPERVISED LEARNING FOR UNSTRUCTURED TEXT (NER MODEL WITH SECTION MODELLING)**

**PROCESS** : | Doc | -> | **Spacy NER/Custom GPT3 NER Mode**l| - > \*Name Tuple

- > \*Dates Tuple

-> \*Amount Tuple

**TAKEAWAY : Tuple of All Possible Matches wrt entity**

**STAGE - 2:**

**OBJECTIVE : TRAIN AND CREATE THE BIDIRECTIONAL LSTM BASED ON CONTEXT, LABEL PAIRS, YIELDING US A MULTICLASS CLASSIFICATION FOR NAME MODEL**

**PROCESS : |\***Name Tuple| -> |Window\_length\_Context| -> |**BiDirectional LSTM Model**|

-> |**MULTI CLASS CLASSIFICATION**| -> |Mapped \*\*kwargs of Names >Party One>Party Two

**TAKEAWAY : Party One or Party Two based on Classification**

**STAGE - 3:**

**OBJECTIVE : TRAIN AND CREATE THE BERT MODEL FOR TEXT PROCESSING AND SUMMARIZATION, USING GLOVE SEMENATIC DOC SIMILARITY YIELDING US A MULTICLASS CLASSIFICATION FOR DATE MODEL**

**PROCESS : |\*D**ate Tuple| -> |Doc| - > |**BERT Text Processing|(Lowers the Dimension Takes out irrelevant information keeps entity intact - BERT Text Summarization|** -> |Glove Semantic Similarity| -> |MULTI CLASS CLASSIFICATION| -> |Mapped \*\*kwargs of Dates>Start Date> End Date> Renewal Days

**TAKEAWAY : Start Date, End Date Renewal Days based on Classification**

**STAGE - 4:**

**OBJECTIVE : TRAIN AND CREATE THE BIDIRECTIONAL LSTM BASED ON CONTEXT, LABEL PAIRS, YIELDING US A BINARY CLASS CLASSIFICATION FOR VALUE MODEL**

**PROCESS**: **|\***Amount Tuple| -> |Window\_length\_Context| -> |**BiDirectional LSTM Model**|

-> **|BINARY CLASS CLASSIFICATION**| -> |Mapped \*\*kwargs of Amount - Aggreement\_Amount|

**TAKEAWAY : Agreement Value**

(We Finally Get the Extracted MetaData Mapping and create a csv from them)

**APPROACH KPI'S:**

* **USING STATE OF ART MODELS FOR CLASSIFICATION**
* **MODULARIZED SEPARATE MODEL FOR BETTER EFFICIENCY**
* **MODELS LEARN FROM RELATIVELY LESS DATA**
* **SEMI-SUPERVISED LEARNING FOR UNSTRUCTURED TEXT**

**APPROACH ARCHITECTURE:**

**|DOC|>|NER MODEL|>|BiDirectional LSTM|>|MULTICLASS|> \*\*kwargs**

**|SECTION MODELLING| >|BERT|>|GLOVESEMENATIC SIMILARITY|> \*\*kwargs**

**>|BiDirectional LSTM|>|BINARYCLASS|> \*\*kwargs**

**APPROACH BRIEF:**

* As studying the Agreement we notice that metadata can easily be figured out by using a NER Model(For Best efficiency I used Spacy NER, still a slight better performer is GPT 3 NER)
* By Getting the NER and it's Label we try to figure out by a **SECTION MODELLING the label's context**
* **And then** passing the context and label to feed out Bi Directional model
* Except When we Try to do Date Processing Things become way unstructured because Start Date and End Date Context Usually Matches and are Generally of Some Vicinity Their is a strong chance our model will **OVERFIT To avoid it** We pass out DOC > to a BERT SUMMAZRIZATION| and from their Calculating Semenatic Similairty and the best Match would yield better and much more efficient solution
* Our Bi Directional LSTM will classify the given new text on class it most likely is to Belong
* We Will Have Three Trainig Model
* For the final processing we explode all \*args pair into a df,