



GSA AI/ML EULA Challenge 2020

Karsun Solutions LLC

Problem statement

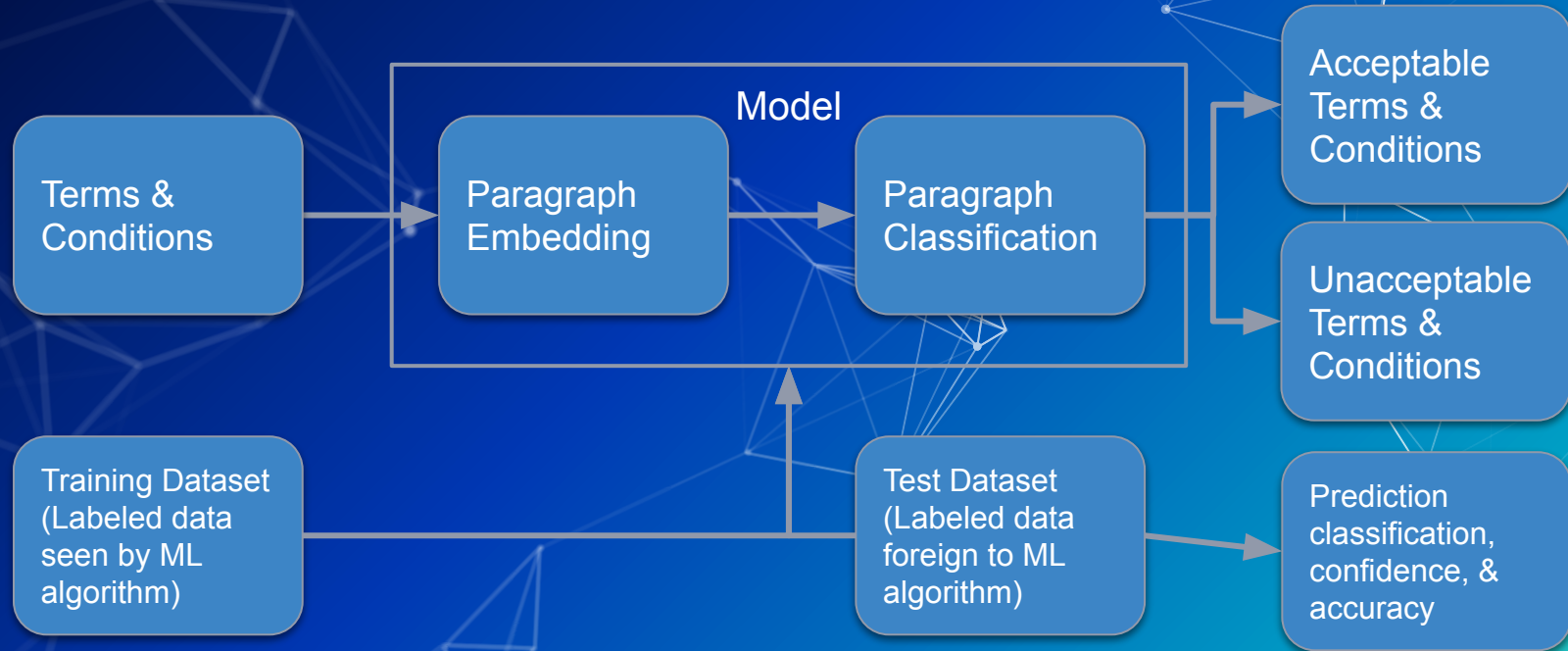
To create an AI/ML based solution/aid to review applicable end-user license agreements (EULAs) and flag the terms and conditions that are not in accordance with Federal law and regulations.

This word cloud visualization represents the semantic structure of the English language. The words are arranged in a network where the size and position of each word reflect its frequency and its relationships to other words. The color-coding indicates the frequency of the words: red for high-frequency/core words, blue for medium-frequency words, and green for low-frequency/niche words. The layout is dense and interconnected, with many words appearing in multiple clusters. The background is a light blue gradient with a subtle grid pattern.

Key clusters and words include:

- Core Words (Red):** The, and, a, of, in, to, that, on, it, he, she, he, was, were, had, do, he, she, he, was, were, had, do, he, she, he, was, were, had, do.
- Medium-Frequency Words (Blue):** The, and, a, of, in, to, that, on, it, he, she, he, was, were, had, do, he, she, he, was, were, had, do, he, she, he, was, were, had, do.
- Low-Frequency Words (Green):** The, and, a, of, in, to, that, on, it, he, she, he, was, were, had, do, he, she, he, was, were, had, do, he, she, he, was, were, had, do.

High level approach based on our intuition



Our Solution Components

UI demo

Model

Architecture considerations

Submitted model architecture

Its accuracy

UI demo

Understanding of the goals of the user

- allow user to provide a EULA documentation in MS Word or PDF format

- successfully parses documents into individual clauses for evaluation

- For each individual clause, the solution reports

 - A classification of acceptable or unacceptable

 - A prediction probability

The user interface of the solution is realistic and usable by a business user

Collecting user review of solution output

Model architecture considerations

Paragraph Embedding

- Average the Word embeddings in the paragraph

- Universal Sentence Encoder / InferSent

- Huggingface Transformers: plug and play various pretrained models

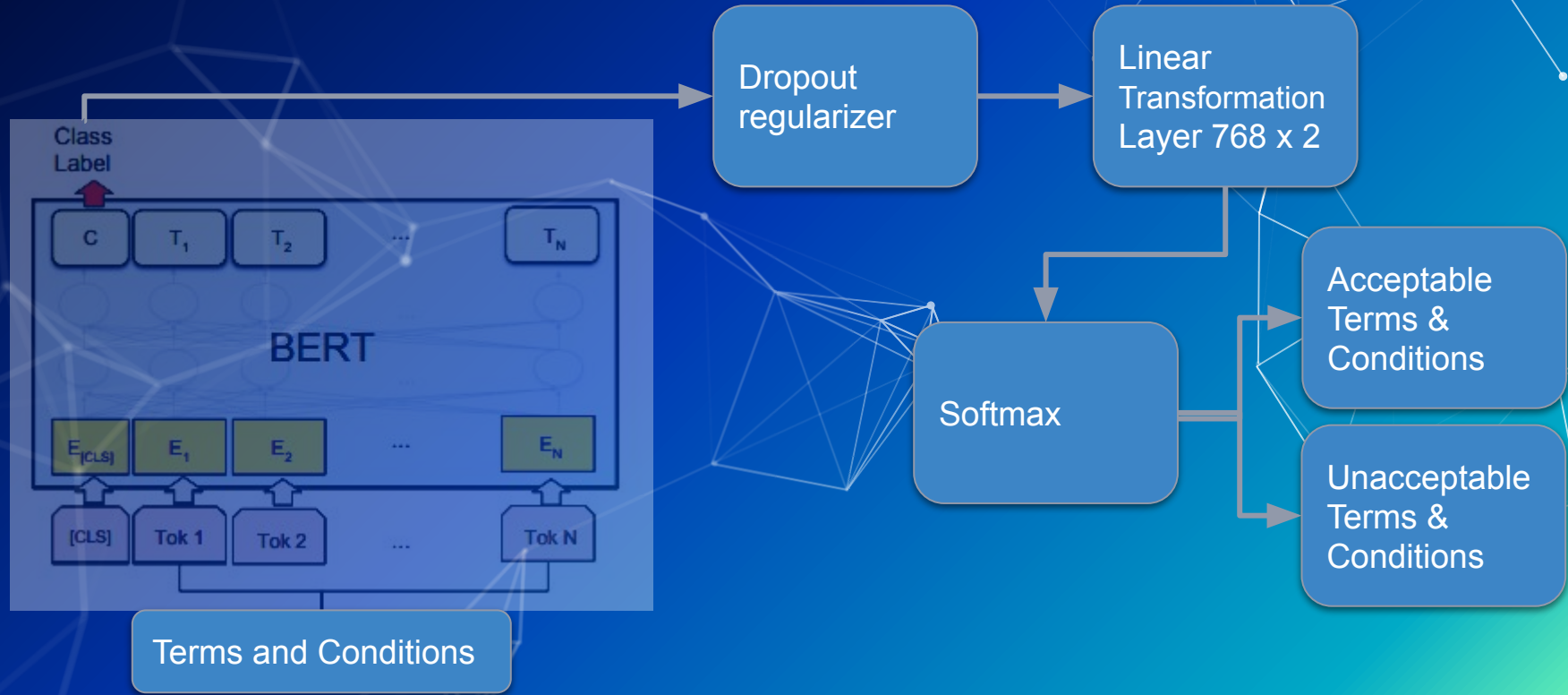
 - Including various versions of BERT, GPT-2, T5, etc.

Paragraph Classification

- k-nearest neighbor

- logistic regression / softmax

Our Model architecture



Accuracy

Test Accuracy: 0.8615 is better than the baseline 0.8132 (when everything acceptable)

Brier Score : 0.1190

Confusion Matrix

	precision	recal	F1
Acceptable	0.91	0.93	0.92
Unacceptable	0.57	0.47	0.52

True\Predicted	Acceptable	Unacceptable
Acceptable	182	13
Unacceptable	19	17

The background is a solid blue gradient that transitions from a darker blue at the top to a lighter blue at the bottom. Overlaid on this are several abstract geometric patterns. These consist of thin white lines connecting small white dots. The dots are positioned at the vertices of various polygons, including triangles, quadrilaterals, and larger, more complex shapes. Some of these shapes are interconnected, creating a network-like structure. The overall effect is a modern, tech-oriented aesthetic.

Thanks