



End-to-End Natural Language Processing on TELUS Business Case : Contract Risk Scoring Model

3666 – Applied Natural Language Processing
School of Continuing Studies, University of Toronto

Presented by :

Mona Hasanzadeh
Ashish Gupta

Let the Journey Begin!



Business Case.



Data Source and Visualizations.



Data Processing and Transformation.



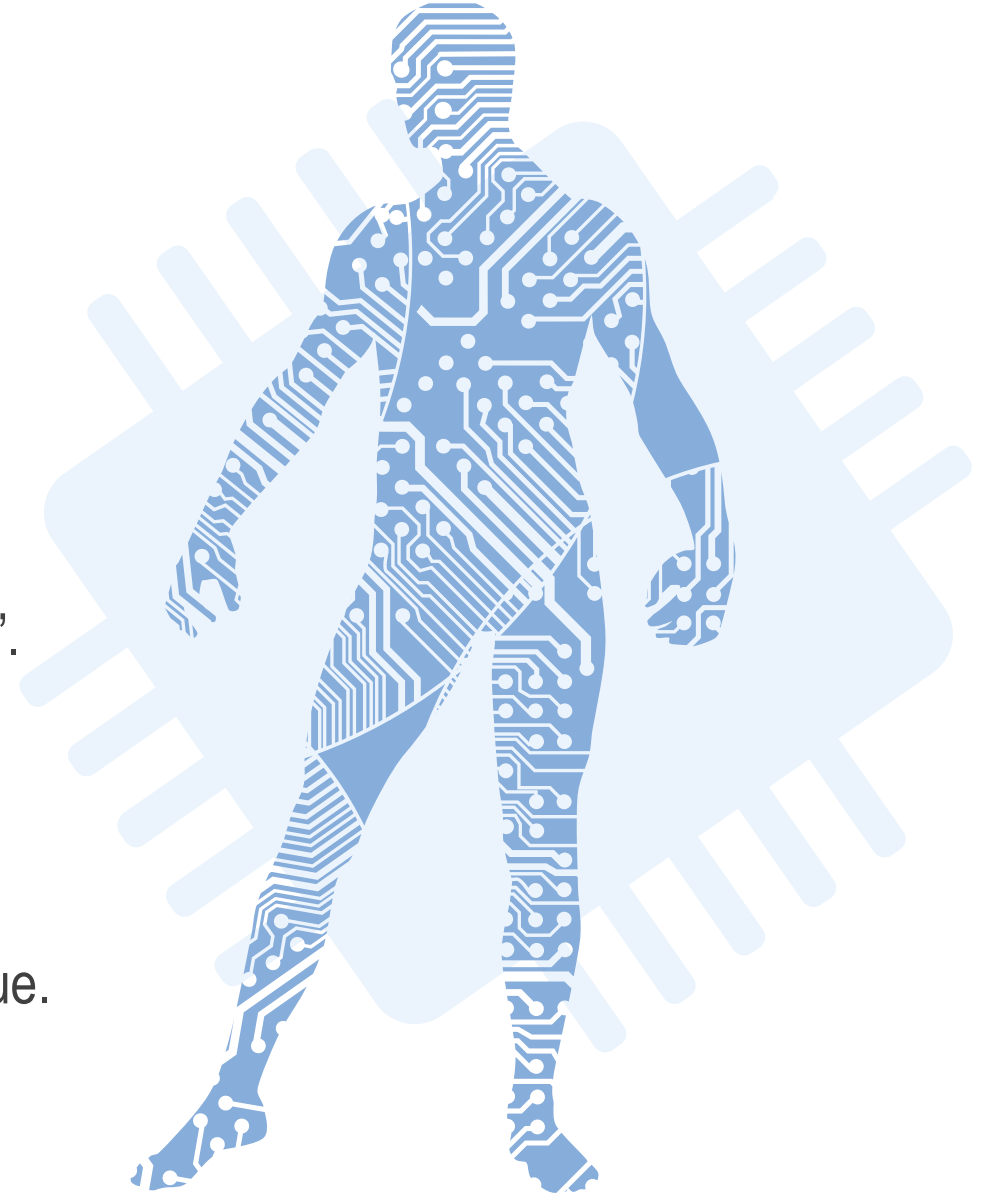
Summarize “Service Level Agreement” and “Warranty”.



Name Entity Recognition to identify Monetary Tags.



Final Predictions : Using K-Means Clustering Technique.



B U S I N E S S

C A S E



Looking at the textual contents of the contract with TELUS suppliers/vendors, identify the risk in each contract and label the contract based on their Risk level using Unsupervised Algorithm.



Extract the key items : “SLA”, “Warranty” to assess and quantify risks for each contract.



Extract the value of the committed funds in all currencies such as “\$”, “£”, “USD” etc.



DATA SOURCE VISUALIZATIONS



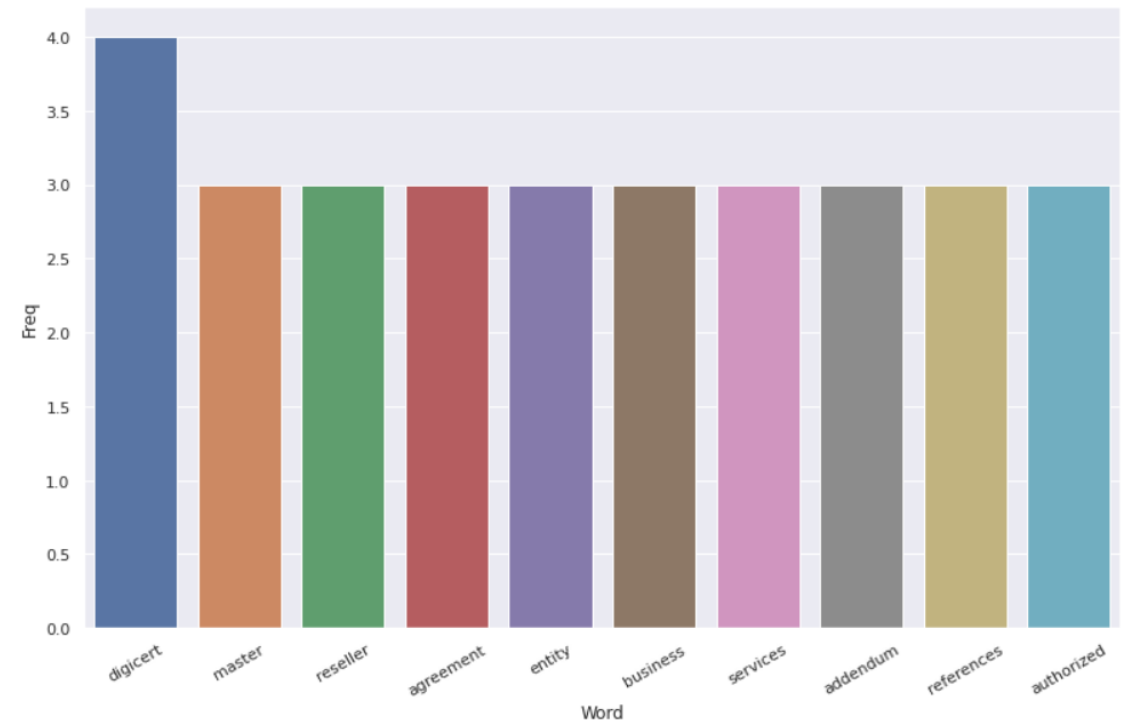
Corpus for Contract Agreements : 21 PDF files

<https://drive.google.com/drive/folders/1uVzyEcpjyDw-gbE-jsZd6mn4Lm5AGNO2>



Corpus Categories

- Master Services Agreement
- Master Reseller Agreement
- Master SaaS Agreement
- Telecom Agreements



DATA PROCESSING TRANSFORMATION



Extract text from PDF, convert it to JSON format and preprocess JSON files and create Pickle objects.



Apply Word Net Lemmatization, POS Tagging via Text Normalizer function.



Extract Key Phrases from the corpus i.e. “Service Level Agreement”, “Warranty”.



Perform Name Entity Recognition to identify Money Entities (“\$”, “£”, “USD”) in the corpus.



Apply TFIDF Vectorization on Key Phrases.

KEY PHRASE EXTRACTION



Text Normalizer function can't extract "Service Level Agreement".



Applied Key Phrase Extraction and Search for "Service Level Agreement".



Extract the index for Starting and Ending paragraph Index.



Generate the Summary from the Extracted Paragraph via a Customized NLTK Summarizer using Similarity Matrix and ranking across sentences.

TEXT SUMMARIZATION



The Summarized Text for Service level Agreement in 6th document is:

In the event of a service degradation or service outage that both impacts your Pilot internet connection and is eligible for an account credit, Pilot will consider those dependent services as affected services for the purposes of credit calculation. Credit Issuance and Refund Policy Pilot will issue any applicable account credits within 15 days of the SLA event. Pilot may conduct such maintenance at any time provided that Pilot has provided Customer with written notice of such maintenance when possible. Credits are issued at the sole discretion of Pilot. The determination of service availability is made at Pilot s sole discretion.

TEXT SUMMARIZATION



The Summarized Text for Service level Agreement in 13th document is:

Items not covered by Support and Maintenance Services InteleX is not obligated to provide Support and Maintenance Services for errors or problems caused by the following third-party components not provided by InteleX, including Customer s infrastructure and network ii use of the SaaS other than in a recommended environment described in the Platform Support Policy available as may be amended from time to time ii training or walk-throughs of the SaaS, change requests, end user requests or troubleshooting with end users or iii continued use of a version of the SaaS for which Support and Maintenance Services is not provided in accordance with Section 3.3 above. Support and Maintenance Services means i online support to Customer s system administrator users relating to technical issues, errors or problems with the SaaS ii access to online resources via the online InteleX customer portal iii notification of and access to InteleX release patches and Documentation released by InteleX and iv notification of and access to Updates and Upgrades of the SaaS

TEXT SUMMARIZATION



The Summarized Text for Warranty in 5th document is:

WITH REGARD TO THE API AND THE TUCOWS SERVICES, TUCOWS AND EACH OF ITS SUPPLIERS A EXPRESSLY DISCLAIMS ALL WARRANTIES AND OR CONDITIONS, EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, TITLE, NON-INFRINGEMENT, THE IMPLIED WARRANTIES AND CONDITIONS OF MERCHANTABILITY OR SATISFACTORY QUALITY AND FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT OF THIRD PARTY RIGHTS B DOES NOT WARRANT THAT THE FUNCTIONS CONTAINED IN THE API OR THE TUCOWS SERVICES WILL MEET THE CUSTOMER S REQUIREMENTS, OR THAT THE OPERATION OF THE API OR THE TUCOWS SERVICES WILL BE UNINTERRUPTED OR ERROR-FREE, OR THAT DEFECTS IN THE API OR TUCOWS SERVICES WILL BE CORRECTED AND C NEITHER WARRANTS NOR MAKES ANY REPRESENTATIONS REGARDING THE USE OR THE RESULTS OF API OR THE TUCOWS SERVICES, OR RELATED DOCUMENTATION IN TERMS OF THEIR CORRECTNESS, ACCURACY, QUALITY, RELIABILITY, OR OTHERWISE. Customer acknowledges and agrees that Tucows does not guarantee that the API or any of the Tucows Services will meet the requirements of Customer or its Users. The API and the Tucows Services are provided as is without any warranty of any kind.

TEXT SUMMARIZATION



The Summarized Text for Warranty in 7th document is:

CUSTOMER'S SOLE AND EXCLUSIVE REMEDY FOR BREACH OF WARRANTY SHALL BE, AT ARISTA'S OPTION, RE-PERFORMANCE OF THE SERVICES, OR TERMINATION OF THE APPLICABLE SERVICES AND RETURN OF THE PORTION OF THE FEES PAID TO ARISTA BY CUSTOMER FOR SUCH NON-CONFORMING SERVICES OR DELIVERABLES. CUSTOMER MUST NOTIFY ARISTA PROMPTLY OF ANY CLAIMED BREACH OF ANY WARRANTIES. "NOTHING IN THIS AGREEMENT SHALL AFFECT THE WARRANTIES PROVIDED WITH ANY HARDWARE PURCHASED OR SOFTWARE LICENSED BY CUSTOMER.

NAME ENTITY RECOGNITION



Extract Committed Funds using search keywords like '\$', 'CAD', 'USD' or 'EUR'.



Extract Paragraph for Committed Funds from a specific document.



Apply Name Entity Recognition for the Paragraph.

NAME ENTITY

RECOGNITION RESULTS



Paragraph : 194 - Committed Funds in "\$" [MRA1.json]

Each Party shall maintain a Commercial General Liability Insurance policy with limits of not less than \$1,000,000 each occurrence, \$2,000,000 general aggregate covering injuries or damage to any person or property which results from their operations or activities under this Agreement.



Money Entities	- {'2,000,000', 'less than \$ 1,000,000'}
Organization Entities	- {'commercial general liability insurance', 'party'}

NAME ENTITY

RECOGNITION RESULTS



Paragraph : 58 - Committed Funds in "\$" [MSA1.json]

5.5 Customer agrees to keep in full force and effect during the term of the Agreement a comprehensive general liability insurance, including contractual liability insurance, in an amount not less than \$1,000,000 per occurrence, providing for the investigation, defense and satisfaction by settlement or otherwise of any claim under the Agreement and b All Risk Property insurance covering all of Customer's personal property located at any of eStruxture's facilities.



Money Entities

- {'less than \$ 1,000,000'}

ADDITIONAL TEXT ANALYTICS METHODS



Extract the Key Entities including “Service Level”, but not SLA.



Extract N – grams.



Generate Part of Speech Tagging for Key Phrases.



Generate a Social Graph and extract entities based on degree centrality and closeness centrality score.

PREDICT RISK

K-MEANS CLUSTERING



Input : TFIDF Vectorization on Key Phrases to access the risk.



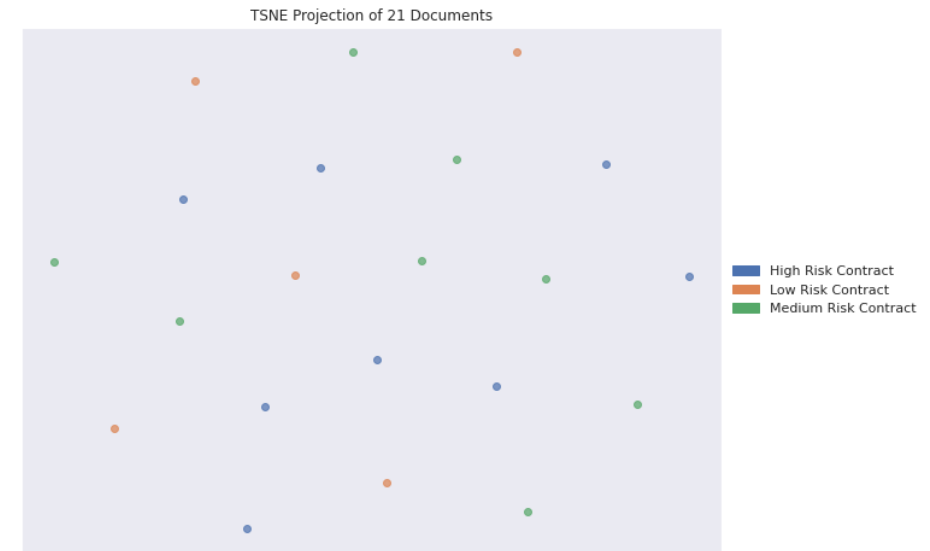
Output : 5 Low, 6 Medium, and 10 High Risk documents.



Initial feedback miss-match our risk assumptions.



More business inputs required from Telus Data Science Team.



SCOPE FOR FUTURE VERSION



Additional measures to cleanse the Text e.g. Replace abbreviation & acronyms based on the nature of business.



Use Pre-Trained word embedding based on business corpus to search for similar key-words.



Extract the text based on the Table of Content automatically.



Apply business knowledge which is required for risk level definition.



Thank You

for bearing us.....