

**RESEARCH TYPE**

24 AUGUST, 2020

**Author**

Ashit Talukder  
ashit.talukder@moodys.com

Daulet Nurmanbetov  
daulet.nurmanbetov@moodys.com

Rupinder Paul Khandpur  
rupinder.khandpur@moodys.com

Albert Aristotle Nanda

Mathew Davis

Chen Li

Sankalp Gaur

**Contact Us**

Americas  
+1.212.553.1658  
clientservices@moodys.com

Europe  
+44.20.7772.5454  
clientservices.emea@moodys.com

Asia (Excluding Japan)  
+852.2916.1121  
clientservices.asia@moodys.com

Japan  
+813.5408.4100  
clientservices.japan@moodys.com

# CoronaPulse: Real-Time Sentiment Analysis and Emergent Multi-Sector Financial Risk Detection From CoVID-19 Events

## ABSTRACT

In this summary, we overview the machine learning system and the components of CoronaPulse. The coronavirus pandemic has had a major, unprecedented impact on the global economy, with some sectors affected much more than others. Due to the fast-moving nature of the epidemic across the globe, it is difficult to understand the multifaceted effect of the pandemic on businesses and economies. To address this challenge, we have developed CoronaPulse - an end-to-end system to detect emergent topics and real-time sentiment associated with the effects of Coronavirus on industry sectors and companies. The system consists of a pipeline of machine learning (ML) and natural language processing (NLP) components that identify relevant articles and progressively filter non-relevant news articles, with advanced language models that detect emerging themes, events, and business-relevant sentiment, to uncover multi categorical risks for global businesses in near-real-time. Our NLP tool for financial risk analytics is available for free at <https://pulse.moodysanalytics.com/>

## KEYWORDS

Machine learning, ML Ops, Natural Language Processing, Sentiment Analysis, Text Analytics, Risk Prediction, Concept Modeling

## 1 INTRODUCTION

CoronaPulse is a novel text analytics system comprised of a number of machine learning models for identifying key risks relevant to a predefined set of companies from real-time news feeds. The text analytics pipeline works on a stream of a massive volume of incoming news articles and progressively filters out non-relevant articles using NLP models at multiple stages. It employs custom NLP models, including Transformer-based models [HuggingFace [n.d.]; Vaswani et al. 2017] that incorporate contextual information to derive semantic meaning from the relevant text information. To enable real-time processing of large amounts of text data, the ML models are deployed on a Kubernetes [Kubernetes [n.d.]] system that allows rapid scaling on-demand to meet real-time SLA requirements. Each individual machine learning module is designed as a micro-service, housed in a Docker [Docker [n.d.]] container, and communication between the components is achieved through a RabbitMQ [RabbitMQ [n.d.]] based messaging service. This allows ML models to be added, reused, or replaced seamlessly as needed. The innovation of our system is in the use of the cutting-edge deep learning-based approaches deployed on a scalable middleware framework that uses language models for Sentiment Analysis (BERT-based tuned on financial data) and uncovers

## Table of Contents

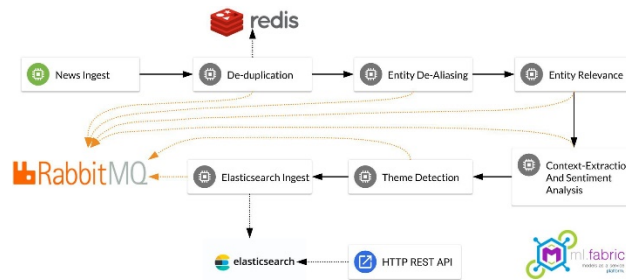
<b>ABSTRACT</b>	<b>1</b>
<b>KEYWORDS</b>	<b>1</b>
<b>1 INTRODUCTION</b>	<b>1</b>
<b>2 TECHNICAL APPROACH</b>	<b>3</b>
2.1 De-duplication	3
2.2 Entity De-aliasing and Resolution	3
2.3 Entity Relevance	3
2.4 Context-Model And Sentiment Analysis	3
2.5 Theme Detection	4
<b>3 RESULTS AND CONCLUSION</b>	<b>4</b>
<b>REFERENCES</b>	<b>5</b>

emerging themes and events (using concept-based models) on relevant information (using NER with a custom model to determine entity relevancy) to provide users unprecedented insights into emergent risks around Coronavirus- related events in near real-time.

## 2 TECHNICAL APPROACH

This section describes individual components used in CoronaPulse as seen in [Fig. 1].

Figure 1 Text analytics pipeline with a series of Machine Learning Modules to detect emerging risk



### 2.1 De-duplication

Newsfeeds are often comprised of duplicative feeds from different sources. We employed a form of locality sensitive hashing, Nilsimsa [Damiani et al. 2004], to identify duplicate articles or near-duplicates among the input article set with minimal computational overhead. Individual hex digests of all incoming articles are stored in a local Redis store [Redis [n.d.]]. Any new incoming articles are first hashed and then compared against the existing hashes, articles with similar hash encodings (less than a "distance threshold") are considered semantically duplicative and rejected.

### 2.2 Entity De-aliasing and Resolution

The de-aliasing module detects unique entities in the set of articles. Companies can be mentioned in multiple forms in the text, including abbreviations, partial mentions, etc. This is further complicated by hierarchical relationships between parent companies and the subsidiaries that need to be resolved. We use an entity normalization function that strips entity names of common words (e.g. such as "corporation" or "group") [Solin [n.d.]], removes punctuation and spacing, and standardizes the entity form. We then match the normalized canonical forms with the normalized list of entities in each article, either by exact match or similarity match within a specific threshold [Navarro [n.d.]]. For resolving company hierarchy, we utilize the Moodys global company ownership data to resolve hierarchies and subsidiaries and normalize to the global parent entity on record (for example, "Ford Credit AG" and "Ford Holdings, Inc." both map to the primary form "Ford Motor Company").

### 2.3 Entity Relevance

News articles and text documents often mention multiple entities, but the article is relevant to only a few of them. Our entity relevance ML module aims to determine relevant entities in the article. A relevant entity, in this context, is taken to be an entity involved in the primary event described in the article. We utilized a custom BERT [Devlin et al. 2018] model which was fine-tuned on a proprietary labeled dataset of finance and risk-related corpus to determine the relevant entities. Candidate entities are first extracted using a CRF-based named entity recognition (NER) model [Honibal and Montani 2017] and then input to the entity relevance module which then classifies them as relevant or not relevant for each entity mentioned in the article.

### 2.4 Context-Model And Sentiment Analysis

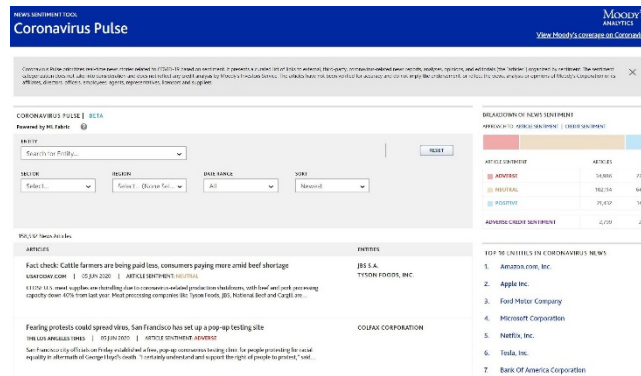
The entity-context and sentiment extraction module extracts the relevant context(s) of the body of an article for a specific target entity(ies) and provides a sentiment score for the article. For each article, the relevant entity forms are discovered in the upstream entity extraction module and for each such relevant entity, contextual information is extracted. We define context as a window of 3 sentences around the entity mention (at relative positions -1, 0, +1).

We utilized a BERT language model [Devlin et al. 2018], fine-tuned on linguistic data in the finance domain, for detecting contextual entity-specific sentiment in the article. The model outputs three logit values for positive/neutral/negative sentiment for each input sentence. The median of all individual sentence sentiments is computed to be representative of article sentiment.

## 2.5 Theme Detection

The theme detection module generates a list of emerging topics based on a collection of articles and also detects those topics in individual articles. This approach uses an unsupervised key-phrase extraction method that utilizes sentence embeddings [Bennani-Smires et al. 2018] and semantic phrase expansion algorithm - sense2vec[Trask et al. 2015] to detect similar semantical concepts from a corpus of text documents. Since it uses phrase/sentence embeddings, the extracted topics are far more superior compared to traditional keyword-based topic models. The Machine-derived topics are further downselected by a human expert to create the final topics for business users of the system.

Figure 2 CoronaPulse is accessible at Moody's Site



## 3 RESULTS AND CONCLUSION

The CoronaPulse text analytics tool and underlying technology demonstrate how a collection of custom machine learning components can be assembled into a pipeline that processes a raw news feed and uncovers risks, themes, and events across multiple sectors. Utilizing transformer-based models[Araci 2019; Hugging- Face [n.d.]] and phrase embeddings (for topic detection) we were able to deliver a production system that achieves high precision in identifying relevant risks. These models in the system are generalizable for many text-analytics applications and can be easily modified for use in other use-cases. Future enhancements could include entity-specific event detection and early warning of sec- tor/corporate credit risk, and broader economic risk by fine-tuning the sentiment models.

## REFERENCES

1. Dogu Araci. 2019. FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. [arXiv:1908.10063](https://arxiv.org/abs/1908.10063) [cs.CL]
2. Kamil Bennani-Smires, Claudiu Musat, Andreea Hossmann, Michael Baeriswyl, and Martin Jaggi. 2018. Simple Unsupervised Keyphrase Extraction using Sentence Embeddings. *Proceedings of the 22nd Conference on Computational Natural Language Learning* (2018). <https://doi.org/10.18653/v1/k18-1022>
3. E. Damiani, S. Vimercati, S. Paraboschi, and P. Samarati. 2004. An Open Digest-based Technique for Spam Detection. *ISCA PDCS 2004*, 559–564.
4. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. [arXiv:1810.04805](https://arxiv.org/abs/1810.04805) [cs.CL]
5. Docker. [n.d.]. *Empowering App Development for Developers: Docker*. <https://www.docker.com>
6. Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. (2017). To appear.
7. HuggingFace. [n.d.]. *Transformers*. <https://huggingface.co/transformers/> Kubernetes. [n.d.]. *Production-Grade Container Orchestration - Kubernetes*. <https://kubernetes.io>
8. Gonzalo Navarro. [n.d.]. A guided tour to approximate string matching. *Comput. Surveys* ([n. d.]).
9. RabbitMQ. [n.d.]. *Messaging that just works - RabbitMQ*. <https://rabbitmq.com>
10. Redis. [n.d.]. <https://redis.io>
11. Paul Solin. [n.d.]. cleanco - clean organization names. <https://github.com/psolin/cleanco>
12. Andrew Trask, Phil Michalak, and John Liu. 2015. sense2vec - A Fast and Accurate Method for Word Sense Disambiguation In Neural Word Embeddings. [arXiv:1511.06388](https://arxiv.org/abs/1511.06388) [cs.CL]
13. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. [arXiv:1706.03762](https://arxiv.org/abs/1706.03762) [cs.CL]

© 2020 Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "MOODY'S"). All rights reserved.

CREDIT RATINGS ISSUED BY MOODY'S INVESTORS SERVICE, INC. AND/OR ITS CREDIT RATINGS AFFILIATES ARE MOODY'S CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES, AND MATERIALS, PRODUCTS, SERVICES AND INFORMATION PUBLISHED BY MOODY'S (COLLECTIVELY, "PUBLICATIONS") MAY INCLUDE SUCH CURRENT OPINIONS. MOODY'S INVESTORS SERVICE DEFINES CREDIT RISK AS THE RISK THAT AN ENTITY MAY NOT MEET ITS CONTRACTUAL FINANCIAL OBLIGATIONS AS THEY COME DUE AND ANY ESTIMATED FINANCIAL LOSS IN THE EVENT OF DEFAULT OR IMPAIRMENT. SEE MOODY'S RATING SYMBOLS AND DEFINITIONS PUBLICATION FOR INFORMATION ON THE TYPES OF CONTRACTUAL FINANCIAL OBLIGATIONS ADDRESSED BY MOODY'S INVESTORS SERVICE CREDIT RATINGS. CREDIT RATINGS DO NOT ADDRESS ANY OTHER RISK, INCLUDING BUT NOT LIMITED TO: LIQUIDITY RISK, MARKET VALUE RISK, OR PRICE VOLATILITY. CREDIT RATINGS, NON-CREDIT ASSESSMENTS ("ASSESSMENTS"), AND OTHER OPINIONS INCLUDED IN MOODY'S PUBLICATIONS ARE NOT STATEMENTS OF CURRENT OR HISTORICAL FACT. MOODY'S PUBLICATIONS MAY ALSO INCLUDE QUANTITATIVE MODEL-BASED ESTIMATES OF CREDIT RISK AND RELATED OPINIONS OR COMMENTARY PUBLISHED BY MOODY'S ANALYTICS, INC. AND/OR ITS AFFILIATES. MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND PUBLICATIONS DO NOT CONSTITUTE OR PROVIDE INVESTMENT OR FINANCIAL ADVICE, AND MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND PUBLICATIONS ARE NOT AND DO NOT PROVIDE RECOMMENDATIONS TO PURCHASE, SELL, OR HOLD PARTICULAR SECURITIES. MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND PUBLICATIONS DO NOT COMMENT ON THE SUITABILITY OF AN INVESTMENT FOR ANY PARTICULAR INVESTOR. MOODY'S ISSUES ITS CREDIT RATINGS, ASSESSMENTS AND OTHER OPINIONS AND PUBLISHES ITS PUBLICATIONS WITH THE EXPECTATION AND UNDERSTANDING THAT EACH INVESTOR WILL, WITH DUE CARE, MAKE ITS OWN STUDY AND EVALUATION OF EACH SECURITY THAT IS UNDER CONSIDERATION FOR PURCHASE, HOLDING, OR SALE.

MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS, AND PUBLICATIONS ARE NOT INTENDED FOR USE BY RETAIL INVESTORS AND IT WOULD BE RECKLESS AND INAPPROPRIATE FOR RETAIL INVESTORS TO USE MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS OR PUBLICATIONS WHEN MAKING AN INVESTMENT DECISION. IF IN DOUBT YOU SHOULD CONTACT YOUR FINANCIAL OR OTHER PROFESSIONAL ADVISER.

ALL INFORMATION CONTAINED HEREIN IS PROTECTED BY LAW, INCLUDING BUT NOT LIMITED TO, COPYRIGHT LAW, AND NONE OF SUCH INFORMATION MAY BE COPIED OR OTHERWISE REPRODUCED, REPACKAGED, FURTHER TRANSMITTED, TRANSFERRED, DISSEMINATED, REDISTRIBUTED OR RESOLD, OR STORED FOR SUBSEQUENT USE FOR ANY SUCH PURPOSE, IN WHOLE OR IN PART, IN ANY FORM OR MANNER OR BY ANY MEANS WHATSOEVER, BY ANY PERSON WITHOUT MOODY'S PRIOR WRITTEN CONSENT.

MOODY'S CREDIT RATINGS, ASSESSMENTS, OTHER OPINIONS AND PUBLICATIONS ARE NOT INTENDED FOR USE BY ANY PERSON AS A BENCHMARK AS THAT TERM IS DEFINED FOR REGULATORY PURPOSES AND MUST NOT BE USED IN ANY WAY THAT COULD RESULT IN THEM BEING CONSIDERED A BENCHMARK.

All information contained herein is obtained by MOODY'S from sources believed by it to be accurate and reliable. Because of the possibility of human or mechanical error as well as other factors, however, all information contained herein is provided "AS IS" without warranty of any kind. MOODY'S adopts all necessary measures so that the information it uses in assigning a credit rating is of sufficient quality and from sources MOODY'S considers to be reliable including, when appropriate, independent third-party sources. However, MOODY'S is not an auditor and cannot in every instance independently verify or validate information received in the rating process or in preparing its Publications.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability to any person or entity for any indirect, special, consequential, or incidental losses or damages whatsoever arising from or in connection with the information contained herein or the use of or inability to use any such information, even if MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers is advised in advance of the possibility of such losses or damages, including but not limited to: (a) any loss of present or prospective profits or (b) any loss or damage arising where the relevant financial instrument is not the subject of a particular credit rating assigned by MOODY'S.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability for any direct or compensatory losses or damages caused to any person or entity, including but not limited to by any negligence (but excluding fraud, willful misconduct or any other type of liability that, for the avoidance of doubt, by law cannot be excluded) on the part of, or any contingency within or beyond the control of, MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers, arising from or in connection with the information contained herein or the use of or inability to use any such information.

NO WARRANTY, EXPRESS OR IMPLIED, AS TO THE ACCURACY, TIMELINESS, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OF ANY CREDIT RATING, ASSESSMENT, OTHER OPINION OR INFORMATION IS GIVEN OR MADE BY MOODY'S IN ANY FORM OR MANNER WHATSOEVER.

Moody's Investors Service, Inc., a wholly-owned credit rating agency subsidiary of Moody's Corporation ("MCO"), hereby discloses that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by Moody's Investors Service, Inc. have, prior to assignment of any credit rating, agreed to pay to Moody's Investors Service, Inc. for credit ratings opinions and services rendered by it fees ranging from \$1,000 to approximately \$2,700,000. MCO and Moody's investors Service also maintain policies and procedures to address the independence of Moody's Investors Service credit ratings and credit rating processes. Information regarding certain affiliations that may exist between directors of MCO and rated entities, and between entities who hold credit ratings from Moody's Investors Service and have also publicly reported to the SEC an ownership interest in MCO of more than 5%, is posted annually at [www.moody.com](http://www.moody.com) under the heading "Investor Relations — Corporate Governance — Director and Shareholder Affiliation Policy."

Additional terms for Australia only: Any publication into Australia of this document is pursuant to the Australian Financial Services License of MOODY'S affiliate, Moody's Investors Service Pty Limited ABN 61 003 399 657 AFSL 336969 and/or Moody's Analytics Australia Pty Ltd ABN 94 105 136 972 AFSL 383569 (as applicable). This document is intended to be provided only to "wholesale clients" within the meaning of section 761G of the Corporations Act 2001. By continuing to access this document from within Australia, you represent to MOODY'S that you are, or are accessing the document as a representative of, a "wholesale client" and that neither you nor the entity you represent will directly or indirectly disseminate this document or its contents to "retail clients" within the meaning of section 761G of the Corporations Act 2001. MOODY'S credit rating is an opinion as to the creditworthiness of a debt obligation of the issuer, not on the equity securities of the issuer or any form of security that is available to retail investors.

Additional terms for Japan only: Moody's Japan K.K. ("MJKK") is a wholly-owned credit rating agency subsidiary of Moody's Group Japan G.K., which is wholly-owned by Moody's Overseas Holdings Inc., a wholly-owned subsidiary of MCO. Moody's SF Japan K.K. ("MSFJ") is a wholly-owned credit rating agency subsidiary of MJKK. MSFJ is not a Nationally Recognized Statistical Rating Organization ("NRSRO"). Therefore, credit ratings assigned by MSFJ are Non-NRSRO Credit Ratings. Non-NRSRO Credit Ratings are assigned by an entity that is not a NRSRO and, consequently, the rated obligation will not qualify for certain types of treatment under U.S. laws. MJKK and MSFJ are credit rating agencies registered with the Japan Financial Services Agency and their registration numbers are FSA Commissioner (Ratings) No. 2 and 3 respectively.

MJKK or MSFJ (as applicable) hereby disclose that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by MJKK or MSFJ (as applicable) have, prior to assignment of any credit rating, agreed to pay to MJKK or MSFJ (as applicable) for credit ratings opinions and services rendered by it fees ranging from JPY125,000 to approximately JPY250,000,000.

MJKK and MSFJ also maintain policies and procedures to address Japanese regulatory requirements.