

## 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

# Adverse Media Mining for KYC and ESG Compliance

## ABSTRACT

In recent years, institutions operating in the global market economy face growing risks stemming from non-financial risk factors such as cyber, third-party, and reputational outweighing traditional risks of credit and liquidity. Adverse media or negative news screening is crucial for the identification of such non-financial risks. Typical tools for screening are not real-time, involve manual searches, require labor-intensive monitoring of information sources. Moreover, they are costly processes to maintain up-to-date with complex regulatory requirements and the institution's evolving risk appetite.

In this extended abstract, we present an automated system to conduct both real-time and batch search of adverse media for users' queries (person or organization entities) using news and other open-source, unstructured sources of information. Our scalable, machine-learning driven approach to high-precision, adverse news filtering is based on four perspectives - relevance to risk domains, search query (entity) relevance, adverse sentiment analysis, and risk encoding. With the help of model evaluations and case studies, we summarize the performance of our deployed application.

## 1 INTRODUCTION

In today's uncertain geopolitical and social environment, global institutions face growing challenges to their risk management processes arising from Non-Financial Risk (NFR) factors. These non-financial risks include, but are not limited to, conduct, cyber, country, compliance, third-party, ESG (Environmental, Social and Corporate Governance) risks. Inadequate compliance & screening controls have cost top banking institutions and other non-financial firms millions of dollars in fines between 2018-2019 alone. In the matter of US Bancorp, fined for lax anti-money laundering controls in 2018 [7], it agreed to a \$613 million (USD) settlement with US regulators. It had failed to report suspicious banking activities carried out by the long-time customer, Scott Tucker, from 2011 to 2013, owner of several payday lending businesses.

With an increased focus on ESG and other regulatory expectations, institutions have realized the importance of integrating adverse news monitoring into their frameworks for managing NFRs. Adverse Media screening involves the introspection of news and other third-party data sources for potential indicators of negative news associated with an entity (person or company). Adverse media mining makes use of open-source indicators (publicly available information) as essential early warning indicators. In a recent study [2, 4], researchers found that Wells Fargo's reputation plummeted after regulators announced the bank's financial fraud. However, Glassdoor reviews signaled the bank had a problem with corporate ethics before the fraud was made public.

## Table of Contents

ABSTRACT	1
1 INTRODUCTION	1
2 SYSTEM OVERVIEW	3
3 DISCUSSION AND CONCLUSION	4
REFERENCES	4

**A fast, automated adverse media mining application.** We showcase a system which can scale to high-volume and diverse unstructured data sources, that provides both real-time & batch processing entity searches for negative news.

1. **A high-precision adverse news filtering pipeline.** We develop a novel pipeline to assess the quality of filtered media by its relevance to the risk domains, target entity, and risk attributes (categories and stages of risk).
2. **A searchable database of adverse media profiles.** We propose a representation of risk profiles to characterize better, search and retrieve adverse entities.

## 2 SYSTEM OVERVIEW

We briefly outline the key components of our system as shown Fig. 1 for adverse media mining.

- a. **News Retrieval & Ingest** For each new (entity) user query, we query a lucene-powered news database using full-text search to fetch all news articles containing the entity mention. For every single query, we cache the articles and track the search period over which query was issued. This helps the system track and monitor what new news data must be fetched for repeat user queries. All news data is stored in an ElasticSearch database.
- b. **Distributed Computing Infrastructure** To process a search query, we make use of Celery-based distributed computing system. We have implemented a “data-funelling” pattern of tasks, where each task reduces the number of articles the next task in the pipeline receives. Each task operates on a single article, scheduled using asynchronous work queues. We operate a cluster 18 compute nodes with a master task scheduler that uses an in-memory, key-value data store for book-keeping.

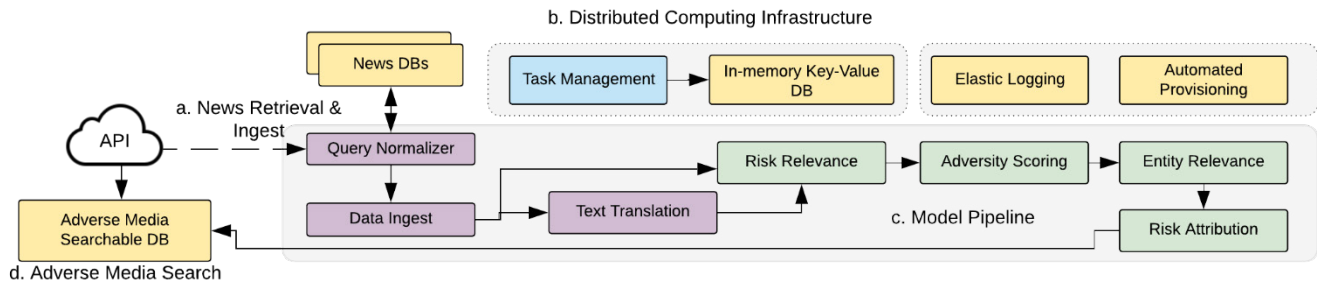


Figure 1: A schematic overview of the adverse media mining system.

- c. **Model Pipeline** There are five primary models in our pipeline.
  - i. **Risk Relevance** - a binary relevance classifier based on supervised training of 2500 articles using a support vector machine model. This model classified the relevance of each article to an in-domain (risk-related) or out-domain class. We achieved 0.81 F1 score over a 80/20 train-test split ratio.
  - (2) **Adversity Scoring** - this is a heuristic model for sentiment scoring that relies on the Loughran-McDonald [6] financial sentiment dictionary for computing adversity score of each article. We further group and weigh differently those sub-sets of keywords that are negative and related to the legal domain.
  - ii. **Entity Relevance** - Each entity (person or an organization) is assumed relevant to an article if it is an apropos risk domain (compliance). We manually tagged 1200 articles to train a supervised binary logistic regression model using bag-of-words features from contexts extracted around the entity mentions using FlairNLP [1] using which we achieved F1 score of 0.8 in 80/20 train-test split.
  - iii. **Risk Categorization** - This step in our pipeline consists of inference risk categories and stage classification. As shown in Fig. 2, we curated a list of fine-grained compliance relevant categories across seven risk types. Using a weakly-supervised labeled data of over 9 Million news documents, we trained CNN-based [5] text classifier for the multi-label classification task. The initial set of 65 categories were expanded using sense2vec [8] to query our internal news database further. Our top-3 categorical accuracy was 0.86 in a 70/30 train-test split.
  - (4) **Risk Stage Identification** - We identified five different criminal proceedings stages that typical compliance events might evolve through. Using a similar weak-supervision approach, we trained a multi-class classifier model using XGBoost [3] with we achieved a F1 score of 0.93 in a 70/30 train-test split.

Figure 2: Distribution of the 65 risk categories across the 7 risk types.

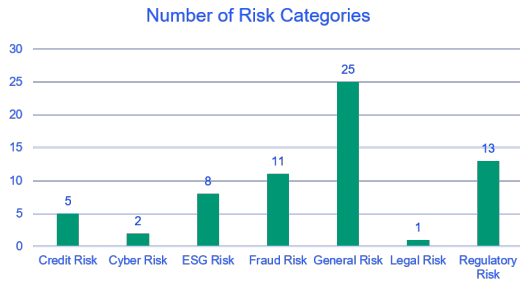
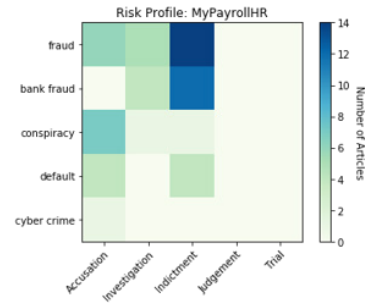


Figure 3: A risk profile for an adverse entity “MyPayrollHR”.



### 3 DISCUSSION AND CONCLUSION

Compliance-related risks are continuously evolving. Adverse news provides a snapshot essentially in time that can help profile such non-financial risk. For example, in Fig. 3, a risk profile for a US- based payroll company MyPayRollHR charged with fraud can be visualized using compliance risk categories and stages with which we can provide actionable insights for the user that best meet their risk appetite. This system is deployed in production in Compliance Catalyst [9], a KYC monitoring tool.

### REFERENCES

1. Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. Contextual String Embed- dings for Sequence Labeling. In *COLING 2018, 27th International Conference on Computational Linguistics*. 1638–1649.
2. Desiree Barry, Sean Brown, Carolyn Ann Geason, Ryan Melehan, Allyson Mac- Donald, Lauren Rosano, and Pedro Henrique Santos. 2019. Measuring Culture in Leading Companies. (2019).
3. Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting sys- tem. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.
4. Yuan Ji, Oded Rozenbaum, and Kyle T Welch. 2017. Corporate culture and financial reporting risk: Looking through the glassdoor. *Available at SSRN 2945745* (2017).
5. Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* (2014).
6. Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66, 1 (2011), 35–65.
7. Pete Schroeder. 2018. U.S. Bancorp to pay \$613 million for money-laundering vio- lations. (2018). <https://www.reuters.com/article/us-usa-usbancorp/u-s-bancorp- to-pay-613-million-for-money-laundering-violations- idUSKCN1FZ1Y>
8. Andrew Trask, Phil Michalak, and John Liu. 2015. sense2vec-a fast and accurate method for word sense disambiguation in neural word embeddings. *arXiv preprint arXiv:1511.06388* (2015).
9. Bureau van Dijk. 2020. *Compliance Catalyst: Risk Management Platform*. <https://www.bvdinfo.com/en-us/our- products/catalyst/compliance-catalyst>

© 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 657AFSL 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.