ISSN 2708-2032
e-ISSN 2708-2040

INTERNATIONAL
UNIVERSITY

INTERNATIONAL JOURNAL OF INFORMATION & COMMUNICATION TECHNOLOGIES

Volume 2, Issue 2
June, 2021

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF EDUCATION AND SCIENCE OF THE REPUBLIC OF KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ УНИВЕРСИТЕТІ МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ
INTERNATIONAL UNIVERSITY

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

Том 2, Выпуск 2
Июнь, 2021

Издание зарегистрировано Министерством информации и общественного развития Республики Казахстан. Свидетельство о постановке на учет № KZ82VPY00020475 от 20.02.2020 г.

Журнал зарегистрирован в Международном центре по регистрации сериальных изданий ISSN (ЮНЕСКО, г. Париж, Франция)

Выходит 4 раза в год.

УЧРЕДИТЕЛЬ:
АО «Международный университет информационных технологий»

ISSN 2708-2032 (print)
ISSN 2708-2040 (online)

СОДЕРЖАНИЕ

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНЖЕНЕРИЯ ЗНАНИЙ

ИНФОКОММУНИКАЦИОННЫЕ СЕТИ И КИБЕРБЕЗОПАСНОСТЬ

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

ЦИФРОВЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И МЕНЕДЖМЕНТЕ

CONTENTS

SOFTWARE DEVELOPMENT AND KNOWLEDGE ENGINEERING

INFORMATION AND COMMUNICATION NETWORKS AND CYBERSECURITY

SMART SYSTEMS

DIGITAL TECHNOLOGIES IN ECONOMICS AND MANAGEMENT

МАЗМҰНЫ

БАҒДАРЛАМАЛЫҚ ҚАМТАМАНЫ ӘЗІРЛЕУ ЖӘНЕ БІЛІМ ИНЖЕНЕРИЯСЫ

АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ЖЕЛІЛЕР ЖӘНЕ КИБЕРҚАУІПСІЗДІК

ИНТЕЛЛЕКТУАЛДЫ ЖҮЙЕЛЕР

ЭКОНОМИКА ЖӘНЕ БАСҚАРУДАҒЫ САНДЫҚ ТЕХНОЛОГИЯЛАР

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ
INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 2. Is. 2. Number 06 (2021). Pp. 75–84
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2021.06.2.010
УДК 530.1, 681.3.06

Azanov N.P.*, Khabirov R.R., Amirov U.E.

Al-Farabi Kazakh National University, Almaty, Kazakhstan

COMPETITIVE INTELLIGENCE AND DECISION-MAKING ALGORITHM USING MACHINE LEARNING FOR INDUSTRIAL SECURITY

Abstract. The purpose of this scientific article is to show what competitor data analytics can do with machine learning and neural networks. In this study, we analyzed data on potential partners of the Department of Defense Office of Hearings and Appeals (DOHA) of the USA and obtained a trained algorithm that can help in making decisions based on keywords, which can minimize reputational risks. The published dataset of the Department of Defense Office of Hearings and Appeals (DOHA) of the USA was selected for analysis of the initial data, which displayed the results of the screening of potential partners along with a text justification. This is the reason why we used Recurrent Neural Network (RNN) instead of Convolutional Neural Network (CNN). Neural networks are a very important part of machine learning. As a result, we have developed a trained machine learning model for recommending the best partners, that is, more proven partners, both professional and reputable. In addition, the developed machine learning model does not allow working with an organization of bad partners who could act in bad faith and carry reputational risks.

Keywords: Competitive Intelligence (CI), Data Analysis, Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Machine Learning (ML)

Introduction

If we talk about competitive intelligence (CI), we can easily find that CI is a systematic collection and analysis of information from multiple sources, as well as a coordinated CI program. This is the act of identifying, collecting, analyzing, and disseminating information about products, customers, competitors, and any aspects of the environment necessary to support managers in making strategic organizational decisions. CI means understanding and learning about what is happening in the world outside of business to improve your competitiveness, i.e. learning as much as possible about your external environment, including the industry as a whole and the relevant competitors.

Key points:

Competitive intelligence is a legal business practice, as opposed to industrial espionage, which is illegal.

The main focus is on the external business environment.

This is the process of collecting information, converting it into intelligence, and then using it in decision-making. Some CI professionals wrongly emphasize that if the information collected is not usable or actionable, it is not intelligence.

Another definition of CI sees it as an organizational function responsible for early identification of risks and opportunities in the market before they become apparent ("early signal analysis"). The term CI is often seen as synonymous with competitor analysis, but competitive intelligence is more than competitor analysis; it encompasses the entire environment and stakeholders.

Dataset & Preparation

Finding the right dataset is not always easy. As mentioned earlier, there are more datasets in today's world that one person can handle. Nevertheless, finding the right one that fits your idea can be challenging. We also ran into some problems while searching for data. The main issue was that some of the datasets hadn't had the features we were looking for. Suffice it to say that such a dataset could not exist without us tweaking it a little bit. And that is exactly what we did. We found a very interesting dataset about Industrial Security Clearances on Kaggle and quickly began the data cleaning process. This data contains dates, case numbers, decisions, and decision summaries for over 20,000 cases submitted for review between late 1996 and early 2016. Industry contractors that work for or with the United States Department of Defense and come into contact with secret or privileged information must submit documents for a background check by the government as a part of their contractual obligations. Any employee who fails to get the necessary clearance will not be allowed to work.

Employees may however appeal their decision; in this case the decision will be reviewed and finalized (or reversed) by the Department of Defense Office of Hearings and Appeals (DOHA). This dataset contains summaries of the deliberations and results of such hearings and provides a window into getting security clearance to work as a defense contractor in the United States.

In this research, we looked for answers to the following questions:

- What percentage of appeals were overturned or upheld?
- What percentage of appeals resulted in favorable decisions?
- What were the reasons for decisions to be overturned?
- Can a model be built to predict decisions based on the decision text i.e., an auto-classifier? [1].

We decided to analyze the criteria for selecting suppliers for defense orders for the US Department of Defense, as well as the possibility of appealing against negative results. After our analysis, conducted with the help of machine learning, we determined by what criteria you can get a favorable outcome and become one of the suppliers of defense orders.

The data consists of 5 columns: Date - date of the appeal hearing; Decision Digest - a free text field describing the keywords of the appeal; Decision Identifier - the type of leadership (e.g. financial, foreign influence, security, etc.) to which the appeal relates; Favorable Decision - a binary variable indicating the outcome of the appeal, i.e. favorable, which is rendered in favor of the party, or unfavorable, which leads to the rejection of the decision on security; Clearance Upheld - a binary variable indicating whether the decision made before the appeal was upheld by the Court of Appeal or rejected. "Yes" indicates that the judge's decision was upheld, "No" indicates that the judge's decision was overturned.

	A	B	C	D	E	F	G
1	**N**	**casenum**	**date**	**digest**	**keywords**	**Favorable decision**	**Decision upheld**
2	0	15-08250.a1	07/28/2017	The Judge's adverse findin	Guideline C; Guideline B	No	Yes
3	1	15-03801.a1	07/28/2017	Applicant argues that the J	Guideline B	No	Yes
4	2	15-07971.a1	07/26/2017	Applicant's appeal brief co	Guideline F	No	Yes
5	3	15-03098.a1	07/26/2017	The Board cannot conside	Guideline F	No	Yes
6	4	14-04693.a1	07/26/2017	Applicant contends that his	Guideline F; Guideline E	No	Yes
7	5	16-00844.a1	07/25/2017	Applicant claims the two u	Guideline F	No	Yes
8	6	15-07009.a1	07/25/2017	Applicant was charged wit	Guideline E	No	No
9	7	15-03162.a1	07/25/2017	The Judge stated that any	Guideline K; Guideline E	No	Yes
10	8	16-00757.a1	07/24/2017	Despite Applicant's argume	Guideline F	No	Yes

Figure 1 – Dataset file preview details

Figure 1 shows the file consisting of decision text from Security Appeal hearings at the US Department of Defense. The decisions are classified by:

- Keyword – the security guideline under which clearance is sought such as Financial, Foreign influence etc.
- Favorable decision - whether the Appeals decision was favorable (security clearance granted) or unfavorable (security clearance rejected).
- Decision upheld - whether the original decision was upheld or not [1].

Analysis (% of Yes or No)

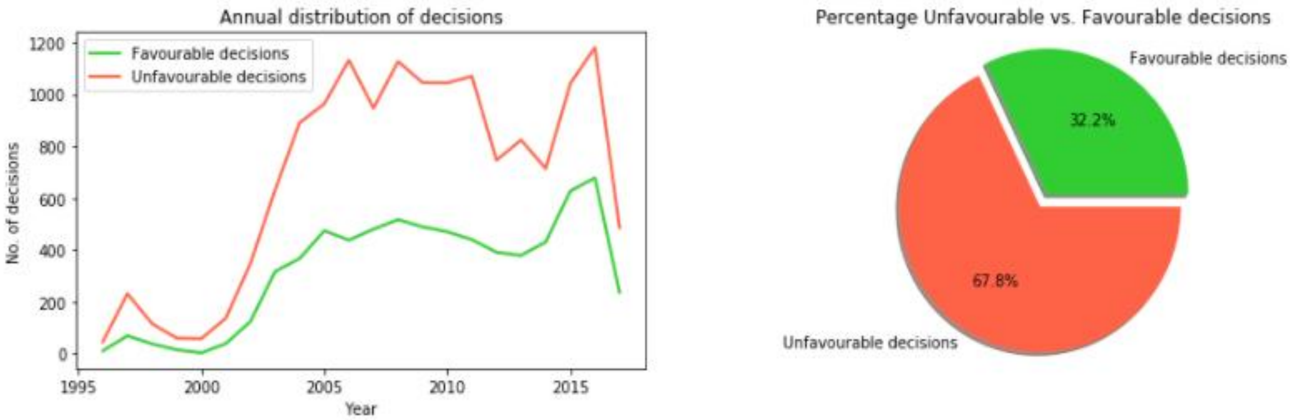

Figure 2 - Dataset analysis

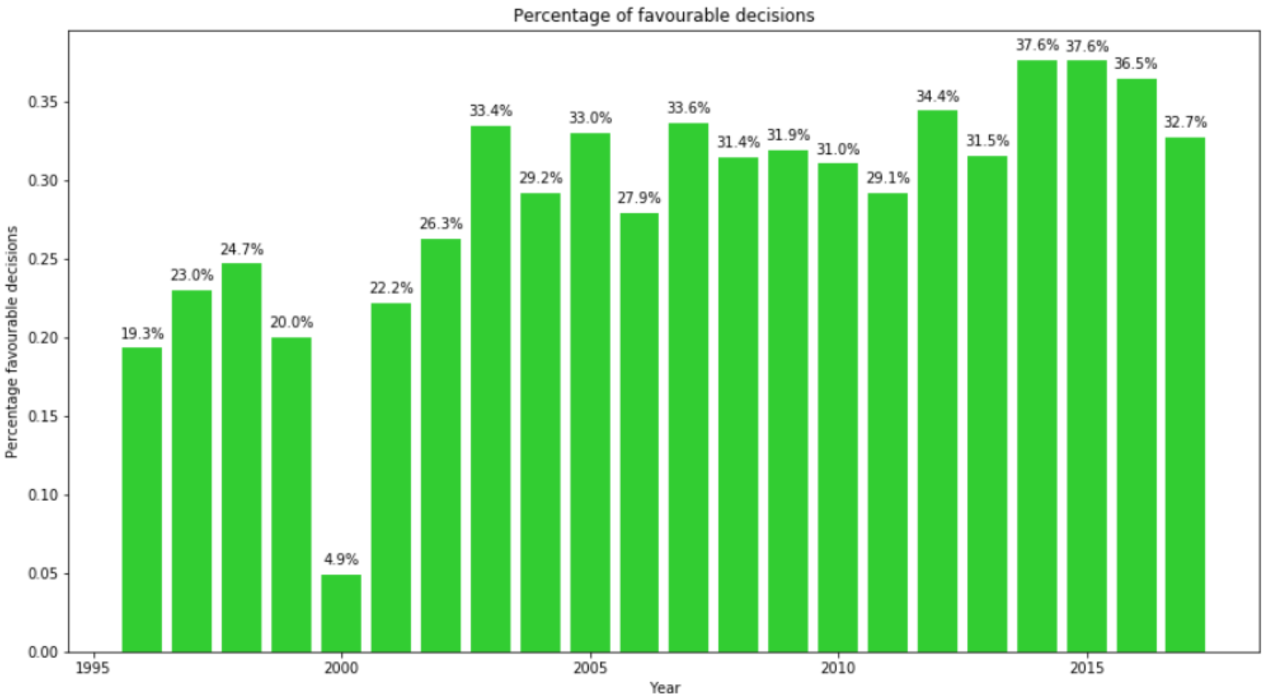

Figure 3 - Percentage of favourable decisions

It's evident that the majority of the cases heard by the Appeals court result in unfavorable decisions i.e. the security clearance is rejected. Almost 68% of the cases heard are rejected, which would reinforce the notion that obtaining a security clearance is a difficult task! Moreover, from the first chart, it appears that the period from 2003 onwards has seen a marked increase in the number of appeals that the court is receiving. We found, that with the increased volume of cases, the rate of rejection has increased, spiking to 37.6% in 2014 and 2015.

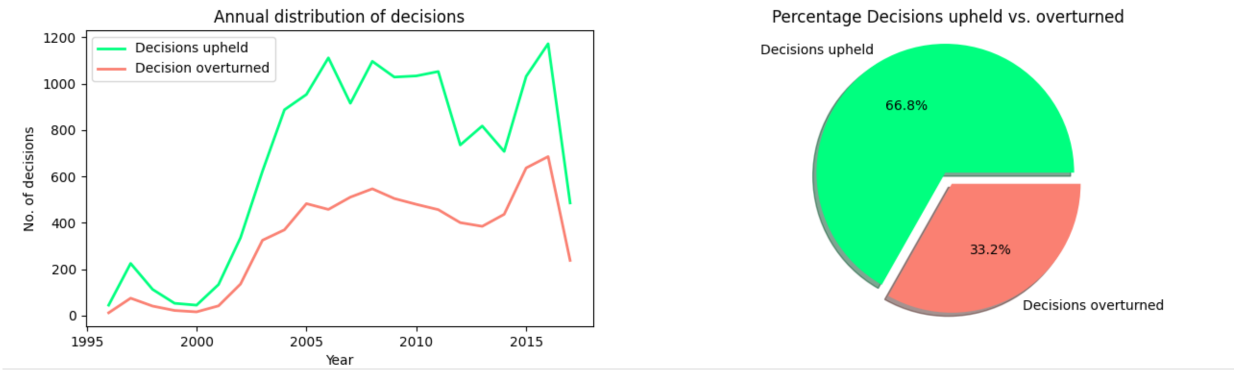

Figure 4 - Statistics about data [1]

Considering the increased volume and rate of rejection, it might be indicative of a more lenient process for appeals, where the court has been receiving more and more cases without merit. The subsequent set of charts looks at exploring the upheld decisions - visualization of the number upheld and overturned decisions by year, the proportion of upheld and overturned decisions, and the percentage of upheld decisions by year. Almost 1 out of 3 cases heard by the Appeals court is upheld. From the first glance at the numbers, it seems like there might be a relation between the types of cases that are upheld and the decision itself (favorable or unfavorable). Decisions overturned spiked in 2014 and 2015 at 38%, also the years in which the Appeals court received its highest volume of cases.

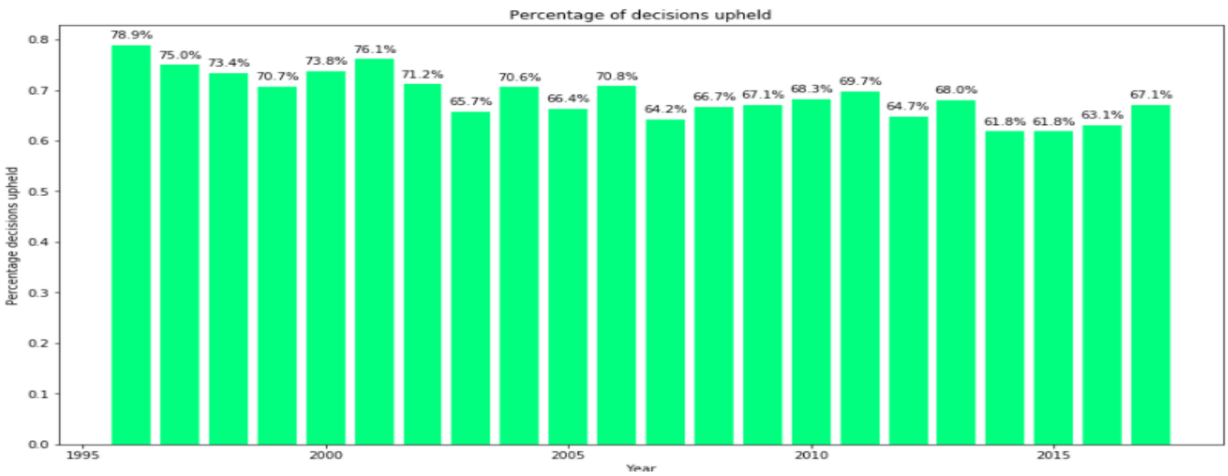

Figure 5 - Percentage of decisions upheld

Decisions overturned spiked in 2014 and 2015 at 38%, also the years in which the Appeals court received its highest volume of cases. It might be worth cross-tabulating overturned decisions with the outcome of the decision (favorable or unfavorable).

ML algorithms & RNN

In this part of the paper, we will discuss the data manipulation process which is the first step after planning that a Data Scientist undertakes. So, we followed good practice and prepared our datasets. Some of the things done include deleting the unwanted features. Some of the features were tedious and removing such redundant data gave us a better view of the important features that we were to use for our algorithm. Then dropping the entries that had missing values or fixing them where it could be done. Unfortunately, some of the entries had a lot of missing values and there was no way to predict them without losing the main goal of our project. We say this because there are a lot of algorithms out there for fixing missing values and finding patterns to predict what those values could be, but for the purpose of simplicity, we just dropped those entries and moved on to the next step.

Neural networks are one of the most popular machine learning algorithms currently. Over time, it has been conclusively proven that neural networks are superior to other algorithms in accuracy and speed. With various options like CNNs (Convolutional Neural Networks), RNNs (Recurrent Neural Networks), AutoEncoders, Deep Learning, and so on, neural networks are slowly becoming to data scientists or machine learning practitioners what linear regression was to statisticians. Thus, it is necessary to have a fundamental understanding of what a neural network is, what it consists of, what its scope and limitations are. This article is an attempt to explain the neural network, starting with its most basic building block, the neuron, and then delving deeper into its most popular options like CNN, RNN, etc.

A neural network that has more than one hidden layer is commonly referred to as a deep neural network. Convolutional neural networks (CNN) are one of the variants of neural networks widely used in the field of computer vision [3]. The name comes from the type of hidden layers it consists of. Hidden CNN layers are usually composed of convolutional layers, merged layers, fully connected layers, and normalization layers. Here it simply means that instead of using the normal activation functions defined above, the fold and merge functions are used as the activation functions.

Recurrent neural networks, or RNNs as they are called in short, are a very important variant of neural networks widely used in natural language processing. In a conventional neural network, the input is processed through several layers, and the output is made with the assumption that two consecutive inputs are independent of each other. This assumption, however, is incorrect in a number of real-world scenarios. For example, if someone wants to predict the price of a stock at a given point in time or wants to predict the next word in a sequence, the dependency on previous observations must be considered. RNNs are called returnable because they perform the same task for each element of the sequence, with an output that depends on previous computations. Another way to think about RNNs is that they have a "memory" that collects information about what has been calculated so far [3]. In theory, RNNs can use information in arbitrarily long sequences, but in practice they are limited to looking at just a few steps. RNNs take full advantage of the context of sequential data and offer an effective and scalable model for several learning problems related to sequential data. RNNs have enabled significant progress in the fields of text processing, speech processing, and machine translation [4]. Network traffic is made up of packets, and the payloads in the packets appear in the form of data streams. When analyzing the semantics of payloads, we employ an RNN model to classify them. RNNs learn feature representations from training data by storing previous states. In the training phase, the RNN model identifies specific sequences that can distinguish normal communications from anomalous communications. RNNs can also directly process data streams; thus, RNN scans support end-to-end attack detection [4].

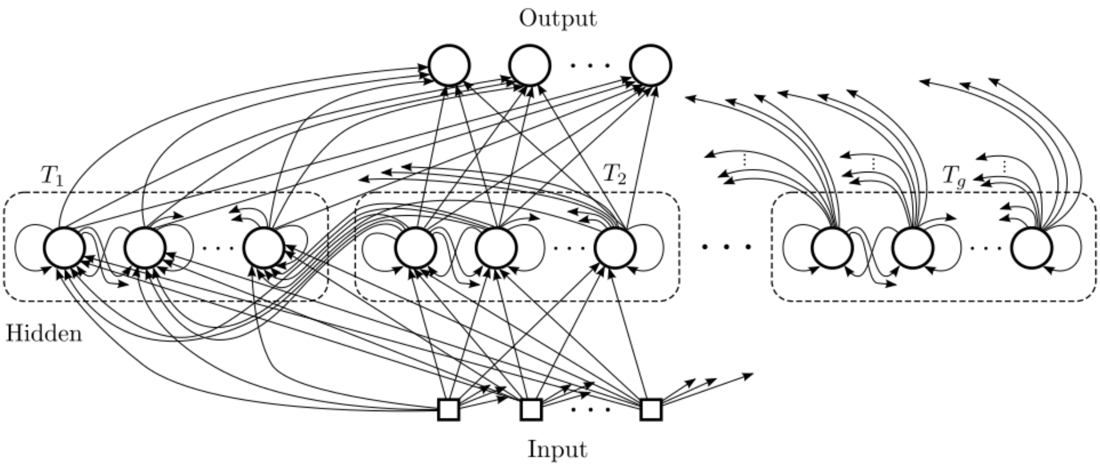

Figure 6 - RNN with an input, output and hidden layer

Figure 6 shows the architectural structure of the Recurrent neural network with an input layer, a couple of hidden layers and an output layer. RNNs have proven to be very successful in natural language processing, especially with their variant of LSTM, which is capable of looking back longer than RNNs.

Implementation of RNN

The data [1] consists of 5 columns: Date - date of the appeals hearing and decision; Digest - free text field describing the appeals decision; Keywords - identifies the type of guideline (such as Financial, Foreign Influence, Security, etc.) that the appeal relates to; Favorable Decision – a binary variable indicating the outcome of appeal i.e. favorable, which is ruled in favor of the party or unfavorable, which results in a rejection of security clearance; Decision Upheld – a binary variable indicating whether the decision made prior to appeal was upheld by the Appeals court or rejected, “Yes” indicates that the judge's decision was upheld, "No" indicates that the judge's decision was overturned; Exploring favorable decisions - visualization of the number of favorable and unfavorable decisions by year, the proportion of favorable and unfavorable decisions, and the percentage of favorable decisions by year.

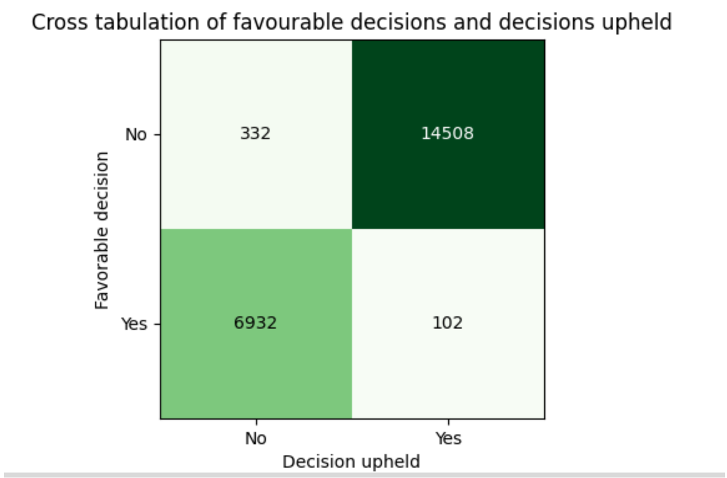

Figure 7 - Statistics about data [1]

Figure 7 shows cross-tabulation which confirms that out of 7,264 (6,932 + 332) decisions overturned, a majority (95% - 6,932 out of 7,264) were initially rejected applications, which were then appealed and overturned in favor of the party. Surprisingly, 5% of the appeals that were initially ruled upon positively were later overturned in the Appeals process and security clearance was rejected. Similarly, out of 14,610 (14,508 + 102) cases that were upheld, 99% were unfavorable (security clearances were rejected). Out of 21,440 decisions that were initially ruled upon unfavorably, 32.3% of such cases (roughly 1 out of 3) were overturned and resulted in a security clearance being granted. Interestingly, an initially favorable decision that goes to the Appeals court is overturned in 76.5% of the cases.

```
EMBEDDING_DIM = 50
model = Sequential()
model.add(Embedding(vocab_size, EMBEDDING_DIM, input_length=max_len))
model.add(Bidirectional(GRU(units=32, dropout=0.15, recurrent_dropout=0.15)))
model.add(Dense(2, activation='sigmoid'))
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
history = model.fit(X_train_pad, Y_train, batch_size=128, epochs=10, validation_data=(X_test_pad, Y_test), verbose=2)
```

Figure 8 - Implementation of Bi-directional RNN with embeddings being trained with Keras

Figure 8 presents a bi-directional RNN that can predict the outcome of whether or not the decision was upheld based on the text of the decision. With a forward RNN model, the accuracy achieved was only in the range of 60% - 70%. Bi-directional RNNs seem apt for this application since the decision text involves several double negatives and indirect references. In our implementation, we have chosen to use “sigmoid” as an activation function, because the final result must be between 0 and 1. In addition, we used “Adam” as an optimizer. Adam is an adaptive torque estimation, another optimization algorithm. It combines both the idea of motion accumulation and the idea of weaker renewal of weights for typical signs. So, at the end of calculations, the result will be demonstrated in accuracy as a metric.

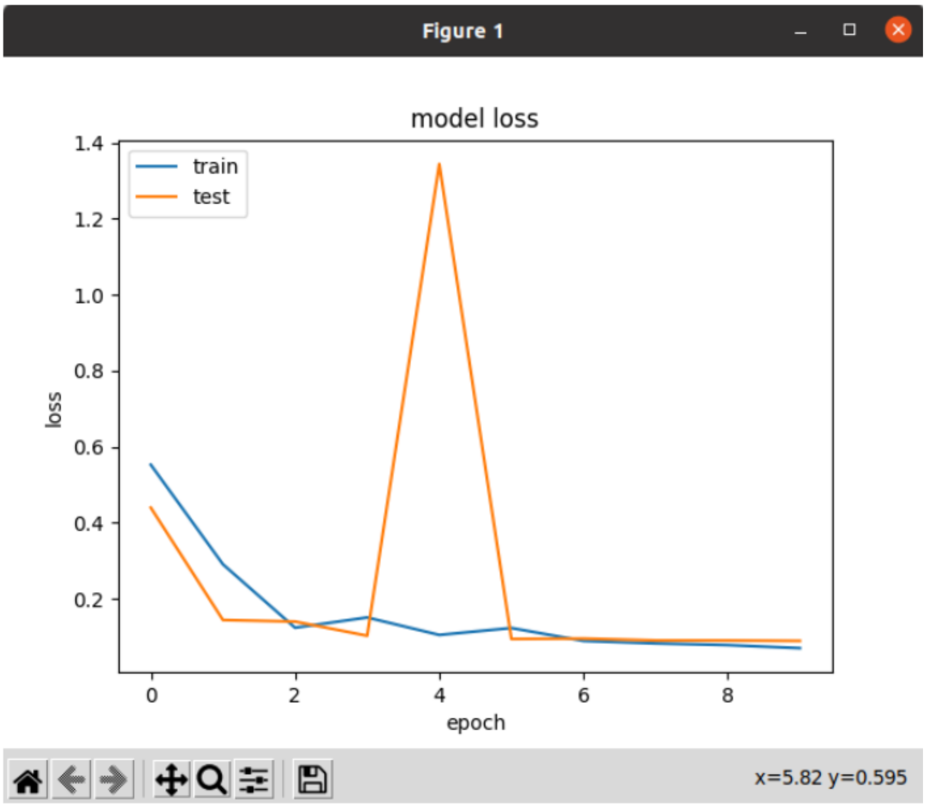

Figure 9 - Graph of result with a comparison train and test datasets by epochs

```
Terminal: Local
2021-04-19 16:25:12.161718: I tensorflow/core/platform/profile_utils/cpu_utils.cc:112] CPU Frequency: 2599990000 Hz
Epoch 1/10
2021-04-19 16:25:15.617940: I tensorflow/stream_executor/platform/default/dso_loader.cc:49] Successfully opened dyna
2021-04-19 16:25:16.286869: I tensorflow/stream_executor/platform/default/dso_loader.cc:49] Successfully opened dyna
137/137 - 205s - loss: 0.5526 - accuracy: 0.7279 - val_loss: 0.4395 - val_accuracy: 0.7899
Epoch 2/10
137/137 - 203s - loss: 0.2905 - accuracy: 0.8885 - val_loss: 0.1439 - val_accuracy: 0.9557
Epoch 3/10
137/137 - 233s - loss: 0.1235 - accuracy: 0.9618 - val_loss: 0.1398 - val_accuracy: 0.9595
Epoch 4/10
137/137 - 159s - loss: 0.1506 - accuracy: 0.9567 - val_loss: 0.1024 - val_accuracy: 0.9710
Epoch 5/10
137/137 - 158s - loss: 0.1045 - accuracy: 0.9717 - val_loss: 1.3445 - val_accuracy: 0.6965
Epoch 6/10
137/137 - 156s - loss: 0.1227 - accuracy: 0.9642 - val_loss: 0.0940 - val_accuracy: 0.9701
Epoch 7/10
137/137 - 156s - loss: 0.0888 - accuracy: 0.9723 - val_loss: 0.0950 - val_accuracy: 0.9694
Epoch 8/10
137/137 - 154s - loss: 0.0822 - accuracy: 0.9738 - val_loss: 0.0900 - val_accuracy: 0.9707
Epoch 9/10
137/137 - 154s - loss: 0.0778 - accuracy: 0.9765 - val_loss: 0.0899 - val_accuracy: 0.9719
Epoch 10/10
137/137 - 153s - loss: 0.0698 - accuracy: 0.9777 - val_loss: 0.0888 - val_accuracy: 0.9719
```

Figure 10 - Results of each epoch in console

Figure 10 shows the result of RRN for 10 epochs with mean **0.8129** accuracy, which is really not a bad result. It can be upgraded and finetuned in the near future by better data preparation with more efficient text analysis algorithms.

Conclusion

As a result, we analyzed data on the refusal or acceptance of commercial partners for cooperation with the Department of Defense Office of Hearings and Appeals (DOHA) of the USA. Classifications of the Decision Outcome (Favorable/Unfavorable decision) and Appeal Outcome (Upheld/Overturned decision) dataset have helped us understand the logical result of the decisions. In this regard, the main data is in the text form, so it was decided to use RNN since Recurrent Neural Networks are better suited for working with text than CNN, that is Convolutional Neural Networks [4].

Finally, we have implemented the RNN ML model with 0.8129 accuracy using the Python Machine Learning tools: numpy, pandas, sklearn, tensorflow and keras. The Recurrent Neural Network Machine Learning Model with high accuracy confirmed the correctness of our research. This RNN ML model can predict answers with not bad accuracy to the question: “Confirm or not confirm commercial partners to work with?” by keywords, digest, and text analysis. We can save this learned RNN ML model to reuse it in next projects or companies. We can apply our solution with this RNN ML model, already trained on this data, to aid decision-making in other ministries, enterprises, companies, etc. Based on the analyzed data, we propose a trained model for weeding out unfavorable partners to Kazakhstan organizations or government agencies.

Since CI is a process associated with collecting information, transforming it into intelligence data, in our case, using our enterprise RNN ML trained model helps commercial and government organizations in decision-making, to defend themselves from unscrupulous partners, espionage, information data leakage, and minimize the reputation risks.

REFERENCES

1. Industrial Security Clearances Classified [Electronic resource] URL: www.kaggle.com/vivek2412/industrial-security-clearances-classified (accessed:15.04.2021)
2. Yang, Jean & Hance, Travis & Austin, Thomas & Solar-Lezama, Armando & Flanagan, Cormac & Chong, Stephen. (2015). End-To-End Policy-Agnostic Security for Database-Backed Applications. [Electronic resource] URL: https://www.ftc.gov/system/files/documents/public_comments/2015/10/00039-97834.pdf

3. Arneric J, Poklepovic T, Aljinovic Z (2014) GARCH based artificial neural networks in forecasting conditional variance of stock returns. Croat Oper Res Rev 5(2):329–343
4. Hongyu Liu, Bo Lang, Ming Liu, Hanbing Yan (2019). CNN and RNN based payload classification methods for attack detection. Knowledge-Based Systems. Volume 163, 1 January 2019, Pages 332-341
5. Junxiang Fan, Qi Li, Junxiong Hou, Xiao Feng, Hamed Karimian, Shaofu Lin (2017). A Spatiotemporal Prediction Framework for Air Pollution Based on Deep RNN. [Electronic resource] URL: https://doi.org/10.5194/isprs-annals-IV-4-W2-15-2017
6. Jan Koutn´ık, Klaus Greff, Faustino Gomez, Jurgen Schmidhuber. A Clockwork RNN. IDSIA, USI&SUPSI, Manno-Lugano, CH-6928, Switzerland. Proceedings of the 31st International Conference on Machine Learning, PMLR 32(2):1863-1871, 2014.
7. Susheel S., Narendra Kumar N.V., Shyamasundar R.K. (2015) Enforcing Secure Data Sharing in Web Application Development Frameworks Like Django Through Information Flow Control. In: Jajoda S., Mazumdar C. (eds) Information Systems Security. ICISS 2015. Lecture Notes in Computer Science, vol 9478. Springer, Cham. [Electronic resource] URL: https://doi.org/10.1007/978-3-319-26961-0_34.

Азанов Н.П. Хабиров Р.Р., Әміров У.Е.

Конкурентная разведка и принятие решений с помощью машинного обучения для обеспечения промышленной безопасности

Аннотация. Цель этой научной статьи показать, на что способна конкурентная разведка и анализ данных с помощью машинного обучения и нейронных сетей. В данном исследовании мы проанализировали данные о потенциальных партнерах Управления слушаний и апелляций Министерства обороны США (ДОХА) и получили обученный алгоритм, который может помочь в принятии решений на основе ключевых слов и который позволяет минимизировать репутационные риски. В качестве анализа исходных данных был выбран опубликованный набор данных Управления слушаний и апелляций Министерства обороны США (ДОХА), в котором наряду с текстовым обоснованием были отображены результаты скрининга потенциальных партнеров. Именно по этой причине мы использовали Рекуррентную нейронную сеть (RNN) вместо Сверточной нейронной сети (CNN). Нейронные сети - очень важная часть машинного обучения. В результате мы разработали обученную модель машинного обучения для рекомендации лучших партнеров, то есть более проверенных партнеров, как профессиональных, так и авторитетных. Кроме того, разработанная модель машинного обучения не позволяет работать организациям с неблагоприятными партнерами, которые могут действовать недобросовестно и нести репутационные риски.

Ключевые слова: Конкурентная разведка, Анализ данных, Рекуррентная нейронная сеть (RNN), Сверточная нейронная сеть (CNN), Машинное обучение.

Азанов Н.П. Хабиров Р.Р., Әміров У.Е.

Өнеркәсіптік қауіпсіздікті қамтамасыз ету үшін машиналық оқытуды қолдана отырып, бәсекеге қабілеттілікті барлау және шешім қабылдау

Аңдатпа. Бұл ғылыми мақаланың мақсаты машиналық оқыту мен нейрондық желілерді қолдана отырып, бәсекеге қабілетті барлау мен деректерді талдауға қабілетті екенін көрсету. Бұл зерттеуде біз АҚШ Қорғаныс министрлігінің (ДОХА) тыңдаулар мен апелляциялар бөлімінің әлеуетті серіктестері туралы деректерді талдадық және кілт сөздерге негізделген шешім қабылдауға көмектесетін және беделді тәуекелдерді азайтуға мүмкіндік беретін оқытылған алгоритм алдық. Бастапқы деректерді талдау ретінде АҚШ Қорғаныс министрлігі (ДОХА) тыңдаулар мен апелляциялар Басқармасының жарияланған мәлімет жиынтығы таңдалды, онда мәтіндік негіздемемен қатар әлеуетті серіктестердің скринингтік нәтижелері көрсетілді. Дәл осы себепті біз конвульсиялық нейрондық желінің (CNN) орнына

қайталанатын нейрондық желіні (RNN) қолдандық. Нейрондық желілер машиналық оқытудың өте маңызды бөлігі болып табылады. Нәтижесінде біз ең таңдаулы серіктестерді, яғни кәсіби және беделді серіктестерді ұсыну үшін машиналық оқытудың оқытылған моделін жасадық. Сонымен қатар машиналық оқытудың дамыған моделі жосықсыз әрекет ете алатын және беделді қауіп-қатерге душар болатын қолайсыз серіктестермен ұйымдарға жұмыс істеуге мүмкіндік бермейді.

Түйінді сөздер: Бәсекеге қабілетті барлау, деректерді талдау, қайталанатын нейрондық желі (RNN), конвульсиялық нейрондық желі (CNN), машиналық оқыту.

Авторлар туралы мәлімет:

Азанов Николай Прокопьевич, физика-математика ғылымдарының кандидаты, әл-Фараби атындағы Қазақ Ұлттық университеті доценті.

Хабиров Роман Рамильевич, әл-Фараби атындағы Қазақ Ұлттық университеті магистранты.

Әміров Уалихан Ержанулы, әл-Фараби атындағы Қазақ Ұлттық университеті магистранты.

Сведения об авторах:

Азанов Николай Прокопьевич, канд. ф.м. наук, доцент, Казахский национальный университет им. Аль-Фараби, Алматы, Казахстан.

Хабиров Роман Рамильевич, магистрант, Казахский национальный университет им. Аль-Фараби.

Әміров Уалихан Ержанулы, магистрант, Казахский национальный университет им. Аль-Фараби.

About the authors:

Nikolai P. Azanov, Cand. Sc. **(**Ph.Math.**)**, Associate Professor, Al-Farabi Kazakh National University.

Roman R. Khabirov, master student, Al-Farabi Kazakh National University.

Amirov E. Ualikhan, master student, Al-Farabi Kazakh National University.

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

Ответственный за выпуск	Есбергенов Досым Бектенович
Редакторы	Далабаева Айсара Касымбековна Джоламанова Балия Джалгасбаевна Медведев Евгений Юрьевич
Компьютерная верстка	Туратауова Айжаркын Ахметовна
Компьютерный дизайн	Туратауова Айжаркын Ахметовна

Подписано в печать 26.06.2021 г.
Тираж 500 экз. Формат 60x84 1/16. Бумага тип.
Уч.-изд.л. 10.1. Заказ №165

Издание Международный университет информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59