sciendo

Optimization of Financial Flows with the Help of Artificial Intelligence in Public Institutions

Daniel IFRIM
The Bucharest University of Economic Studies, Romania
Corresponding author, difrim@gmail.com

Raluca MITULESCU
Corresponding author, mitulescuraluca19@stud.ase.ro

Alecsandra PARNUS[1]
Corresponding author, parnusalecsandra@stud.ase.ro

Eliza GHEORGHE
Corresponding author, gheorghe.eliza@yahoo.com

Abstract. *The only technology that is currently mature enough to help us quickly recover losses in agriculture is digital technology. Digital transformation is increasingly referred to as the fourth industrial revolution, aimed at merging the real world with the virtual one, bringing advantages even in traditional sectors such as agriculture. Artificial intelligence (AI) algorithms are used to analyze large amounts of data collected from various sources in the field, as well as from farmers or beneficiaries of European funds for agricultural sector development. Specifically, they can analyze information gathered from soil sensors, weather stations, satellite images, or European funding applications, among others. This analysis provides farmers with valuable information to make informed decisions regarding seed selection, optimal timing, planting techniques, and resource allocation decisions. At the same time, AI-powered systems can learn from historical data and identify patterns, enabling more accurate predictions and recommendations. Machine learning algorithms can anticipate managerial decisions, detect potential problems, and suggest appropriate interventions to improve financial performance for the successful implementation of projects funded by European funds.*

Keywords: AI, agriculture, application, financials, machine learning.

Introduction

Artificial intelligence (AI), a branch of computer science, refers to systems or machines that mimic human intelligence to perform various repetitive tasks. The term was introduced in 1956 by John McCarthy of the Massachusetts Institute of Technology and encompasses games, expert systems, natural language processing, neural networks, fuzzy systems, genetic algorithms, robotics, and new continuously emerging technologies, indicating the great interest this field enjoys.

The implementation of artificial intelligence systems in institutions enhances productivity efficiency by automating tasks and repetitive processes, allowing staff to focus on more complex and strategic work. It also increases the potential for better decision-making and policy formulation through the analysis of large volumes of data and the identification of patterns and trends from the queried sources, without the need for a programmer's intervention in code creation. In public institutions, this type of software is used for resource planning and the management of operational flows, such as accounting, authorization, and payment processing, as well as procurement

DOI: 10.2478/picbe-2024-0156
© 2024 D. Ifrim; R. Mitulescu; A. Parnus; E. Gheorghe, published by Sciendo.
This work is licensed under the Creative Commons Attribution 4.0 License.

processes carried out in the implementation of projects funded by external funds. Additionally, the entire operational flow from project submission to evaluation, selection, and contracting of European-funded projects are back-office operations managed by the IT system. Although such information systems have often entailed large, costly, and time-consuming implementations with significant investments in hardware or infrastructure, the volume of collected data and cost savings resulting from reduced working hours and operational streamlining have made the investment yield higher returns and have had a tangible impact. The emergence of Cloud-based computing systems and Software as a Service (SaaS) represents one of the main factors influencing the way application implementations are perceived today. The transfer of information systems to the Cloud allows public institutions to simplify their technological requirements, have constant access to innovation, and achieve a faster return on investment.

Literature review

According to Oracle[1], "*traditionally, financial processes such as data entry, collection, verification, consolidation, and reporting have heavily relied on the human factor*." Manual data entry increases production costs, consumes time, and slows down the process of adapting to organizational changes. Many such financial processes, being constant and well-defined, become ideal targets for automation, thereby increasing productivity and efficiency of personnel and accelerating the decision-making process. The emergence of centralized information systems utilizing AI has allowed institutions to centralize and standardize their financial functions. Moreover, according to Oracle[1], "*early automation was rule-based, which means that as a transaction occurred or an input value was entered, it could be subjected to a set of management rules. Although these systems automate financial processes, they require significant manual maintenance, are difficult to update, and lack the flexibility of present-day AI-based automation. Unlike rule-based automation, AI can address more complex scenarios, including complete automation of routine manual processes*." We are witnessing the unstoppable changes brought by the fourth Industrial Revolution in digital technologies. Agriculture is already irrevocably connected to scientific foundations, nanotechnology, satellite technologies, and artificial intelligence. It is not a question of when the agricultural sector will integrate all these technologies but how quickly it will happen. The proliferation of automation in the public-sector management system leads to greater accuracy and correctness in processing the data entered into financial processes. High-volume and routine operations, such as entering data from payment files submitted by beneficiaries, require dedicated personnel subject to fatigue in the process, without excluding human errors or other associated risks. These limitations are overcome with the help of automation within information systems. The processing time for this type of record can be extended to a larger volume of data, and the available staff can be retrained to improve data utilization, enhance European fund absorption procedures, and implement new financial programming periods. Data processed with the help of artificial intelligence over a certain period can generate patterns that can be used to improve decision-making by management teams and financial forecasting. The end result is better data utilization and more time available to the financial team to focus on utilizing this data.

According to a 2021 research report by Savanta and Oracle, 85% of business leaders currently want to leverage artificial intelligence.

By automating manual processes, such as processes related to calculating authorized payment amounts, using the advantages of integrated information systems and artificial intelligence, an analysis of acquired data, identification of key information (e.g., beneficiary name, requested payment amount, financing contract number, installment payment number based on the project), and determining whether or not the beneficiary has any outstanding debts can be achieved. In cases where some of this data is found in other data sources, AI could search for and automatically input them into accounting systems to detect fraud, reconcile accounts, and expedite approvals.

At the same time, automated financial reporting processes to both the European Commission (e.g., ECA - European Court of Auditors) and the Court of Auditors or Audit Authority in Romania enable public institutions to shift employees' activities from manual data collection, consolidation, and reporting to analysis, strategy, and action by planning with greater accuracy the amounts forecasted for payment.

Another advantage of using AI-powered tools by public institutions is the utilization of AI-guided digital assistants that facilitate information retrieval and task completion. For example, institutions managing European funds can use digital assistants to notify accounting teams when expenses are non-compliant or automatically submit payment files for faster reimbursement. Present-day digital assistants are context-dependent, with conversational functionality, and available on almost any device. Employees do not have to memorize a complex query language or transaction codes but only need to run reports for subsequent analysis and expedite the process of European fund allocation.

According to the *"Money and Machines"* report mentioned earlier*, 87% of business leaders believe that organizations that do not rethink their financial processes will face risks, including:*

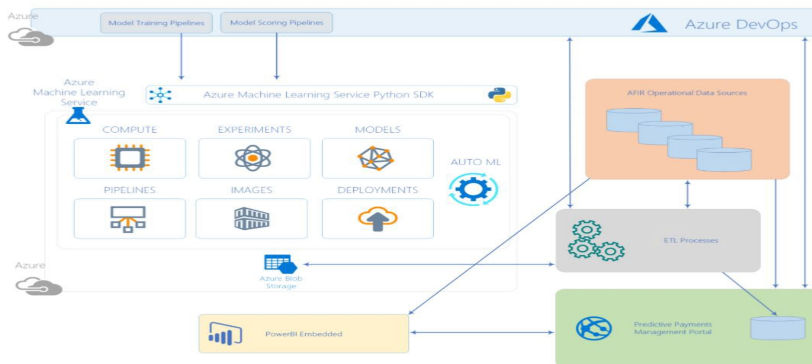
44% regression compared to the competition.
36% more stressed employees
36% inaccurate reporting
35% reduced employee productivity.

Methodology

The model and the areas approached in this paper are: theoretical documentation from specialized articles selectively indicated in the bibliography; practical documentation from own database; data collection and systematization; comparative analysis of the data, interpretation and results. The estimated results for achieving the established objective involve analyzing all data sources related to the information gathered during various stages of project implementation. This includes data retrieval from the evaluation, selection, and contracting stage of the projects, procurement activities, authorization, accounting, and payments.

In this article, we will conduct a detailed analysis of all possible data sources and design a consolidated data model (warehouse-type). We will analyze the attributes of machine learning in detail and identify new attributes. The analysis of data sources for payment prediction will be carried out for a limited number of investment measures within the National Rural Development Program (PNDR), namely: one measure for public beneficiaries, one measure for private beneficiaries with construction-assembly activities, one measure for private beneficiaries without construction-assembly activities, and one measure of type 6.1 or 6.3 from the PNDR. The database contains all the data analyzed during this stage. The components existing at the database level are detailed in Annex 2.

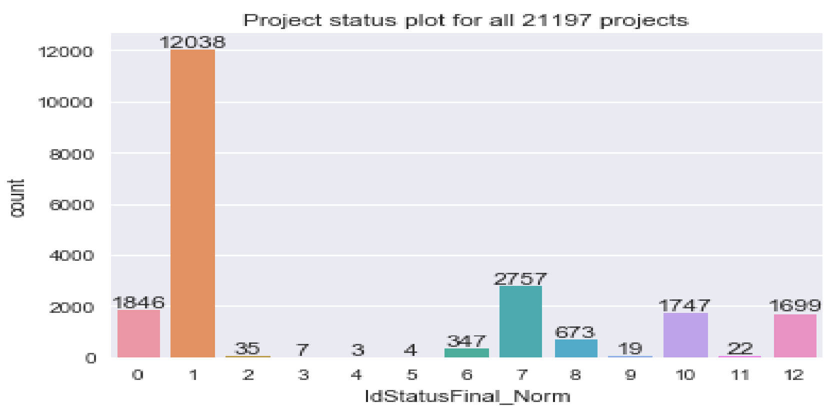
The analysis of the main details of the Proof-of-Concept functional project will be used to predict budgets and their application intervals for reimbursement requests related to 24 different types of projects within AFIR (Agency for Rural Investment Financing). The project is based on approximately 20,000 observations that have been analyzed using Deep Learning technologies, particularly graphs formed by deep recurrent neural networks.



Source: Microsoft Azure Devops.

Data Analysis

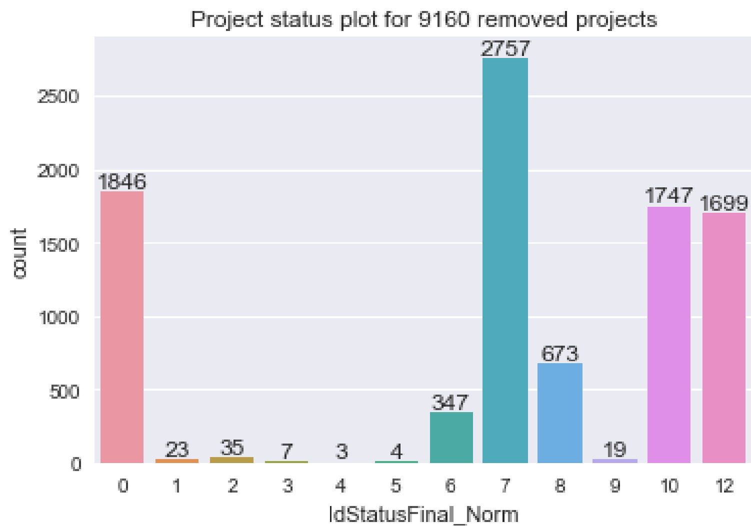
A total of 21,197 project observations were used, with their distribution based on IdStatusFinal_Norm as follows:



Source: The financial accounting system of AFIR.

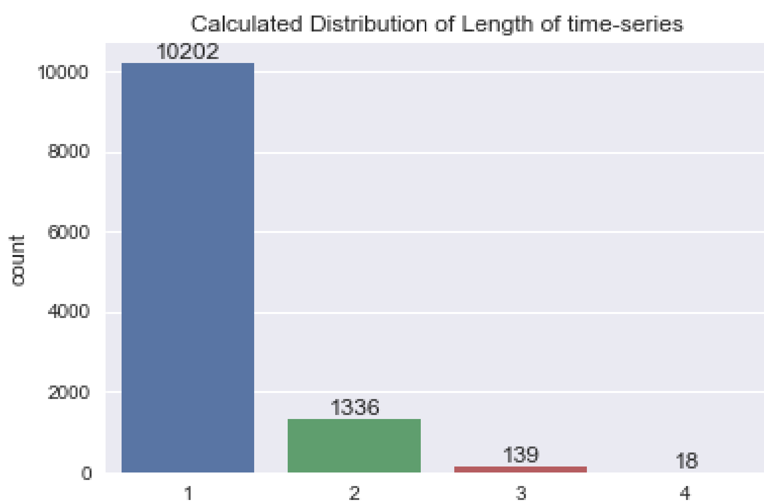
From these basic data, 51 observations were eliminated for which the contracting date was greater than the date of the first reimbursement request, and a total of 9,160 projects had a NULL contracting date, with the following distribution:

DOI: 10.2478/picbe-2024-0156, pp. 1857-1967, ISSN 2558-9652 |
Proceedings of the 18th International Conference on Business Excellence 2024



Source: The financial accounting system of AFIR.

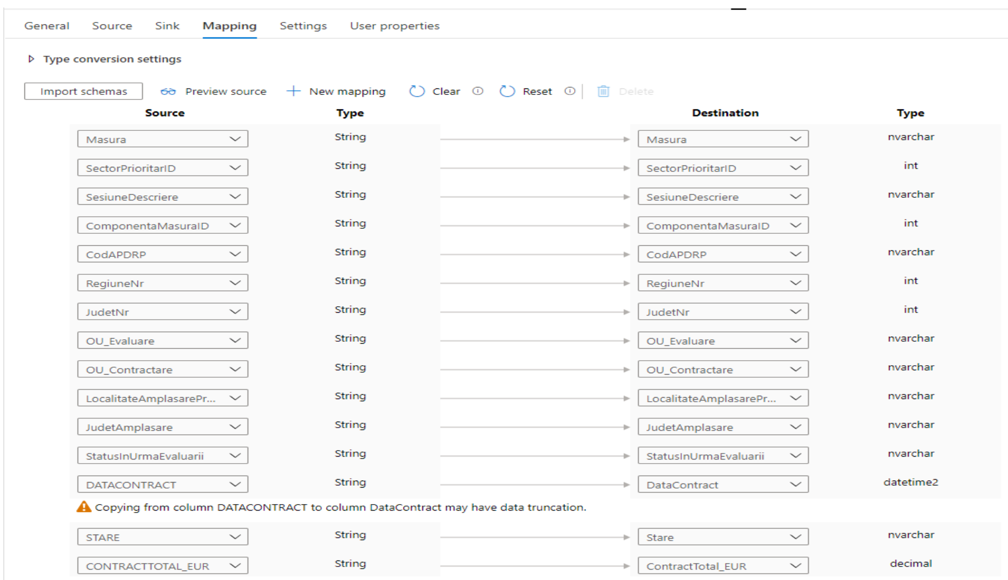
In the end, a working dataset of 11,695 project observations was obtained, which contains the following time series structure:

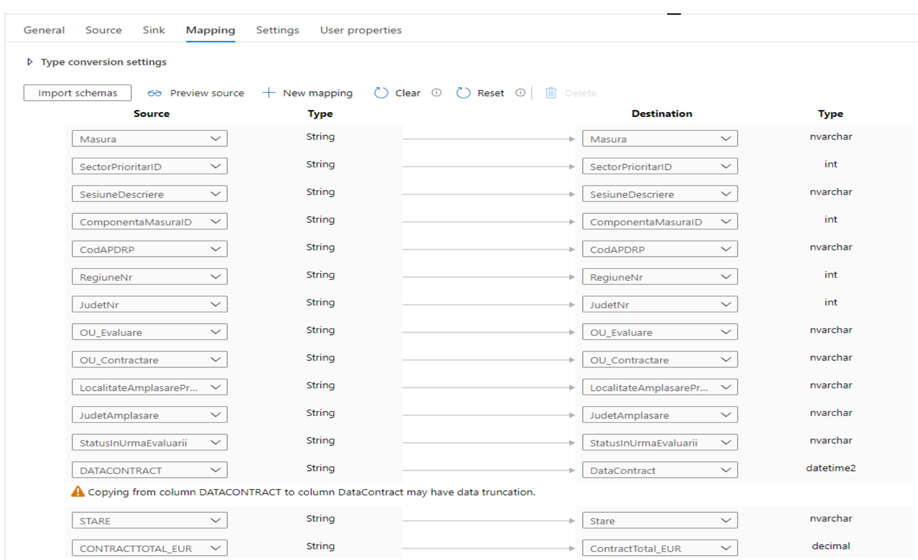


Source: The financial accounting system of AFIR.

As you can see, about 90% of the data contains a single refund request (CR) step, which posed huge problems in the process of training the neural time series prediction model.

The mapping of project properties taken in analysis with tables containing data used in training are detailed in the image below:

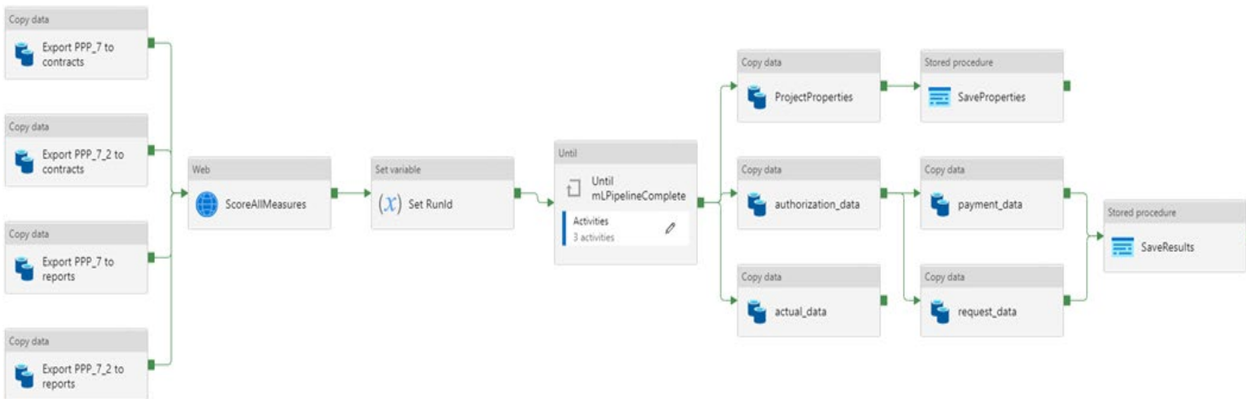




Source: Information from SPCDR (Rural Development Application Processing System of AFIR).

DOI: 10.2478/picbe-2024-0156, pp. 1857-1967, ISSN 2558-9652 |
Proceedings of the 18th International Conference on Business Excellence 2024

The operations to be performed each day (or whenever it is considered necessary to update the training data) are detailed in the graph below:



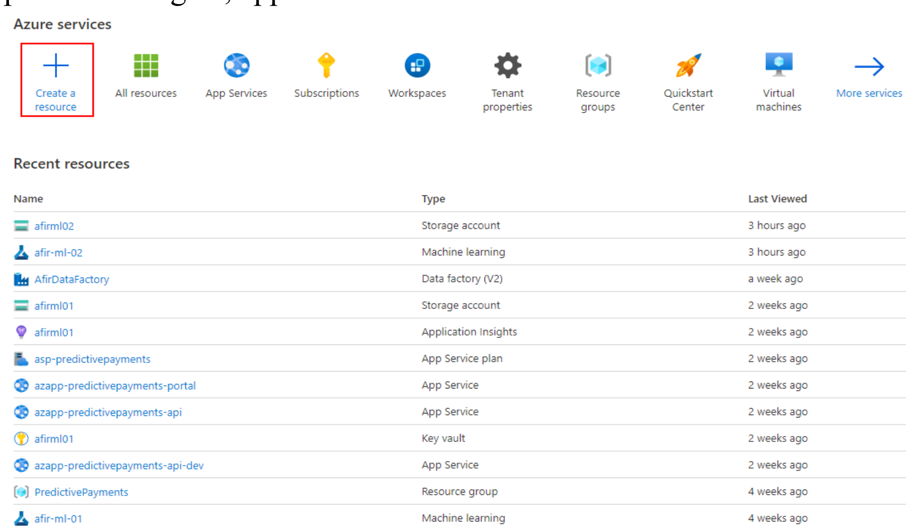
Source: Pilot application carried out at AFIR level.

PoC architecture used: The neural model

The architecture of the neural model of this Proof-of-Concept is based on the principles of deep recurrent acyclic neural graphs. We used the latest research in the field of neural networks with LSTM/GRU memory and in particular their variants optimized for Nvidia GPU processors (CuDNNLSTM). All resources were used from Microsoft Azure infrastructure. The LSTM/GRU architecture is not based on fixed time sequences, being able to both train and differentiate on time series with a virtually unlimited number of steps (there is no limitation to 11 of the number of reimbursement requests). The architecture also uses techniques to skip between the various vertical levels of the graph formed by recurring networks to streamline the learning process (more details in the unfolded sketch of the network below).

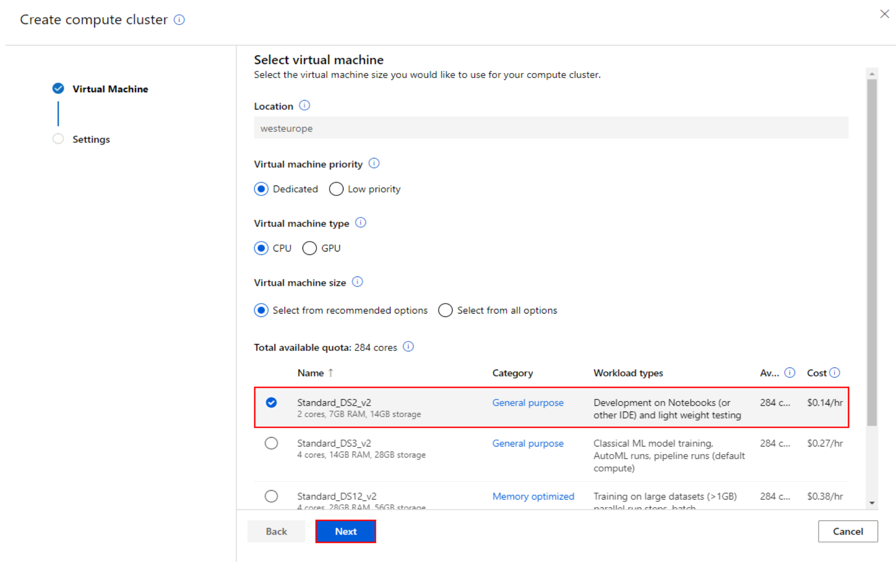
Operationalization

The resources used in Microsoft Azure workspace are storage account, machine learning, data factory, application insights, app services:



Source: Microsoft Azure – pilot project carried out within AFIR

For the cluster used for the question, a standard type DS2_v2 machine was used, which has a very low cost of $ 0.14 / hr.



Source: Microsoft Azure.

The final architecture is based on a variable number of input data divided into two main categories as follows:

- ✓ Continuous variables that are analyzed as floating-point values or as dates.
- ✓ Factor variables that are converted into variables in latent vector spaces (embeddings).

Thus, in the concrete case of the IdCompusNorm variable that captures all types of projects depending on the Sector, Measure and Component of the measure, a semantic (latent) vector space will be obtained in which all types of project with similar behaviors of beneficiaries are "close" within the same cluster.

The main functional modules (object) are:

- AFIRLoader: upload engine and sequential data processing
- AFIRModel: engine for pragmating, training, inference and evaluation of predictive models (based on AFIRLoader)

Minimal Rim

The minimum requirements used in operationalization are:

- Azure instance with GPU processing capability (Nvidia) https://docs.microsoft.com/en-us/azure/virtual- machines/windows/sizes-gpu
- Python 3.5+ installed with Anaconda distribution up to date; (https://blogs.technet.microsoft.com/machinelearning/2017/10/26/anaconda-and-microsoft-partner-to-offer-python-and-r-for-powerful-machine-learning/)
- Minimally installed packages: Keras (https://docs.microsoft.com/en-us/cognitive-toolkit/using-cntk-multigpu- support-with-keras), CNTK

The operationalization system to be used in production.

The production operationalization system of the prediction process consists of the following elements:

Standalone Python script that runs prediction – partially presented at "Script inference."

JSON message (file or otherwise) that configures the request to the system. Presented at "Inference configuration" with the following elements:

a. Define the Models used: "TIME_MODEL" and "VALUE_MODEL."

b. Defines the Data Scaling Methods used. If the input data is already scaled, we will comment on those scaling's (as in the example below)

c. Defines the Input Data for the inference process.

d. Items beginning with "#" are ignored as comments.

Models used as files in H5 format - Hierarchical Data Format (HDF) - two in number whose name is in the JSON message file.

Analysis of results

General results

The results summarized from the training and testing process are presented below:

```
Model:
afir_v4_01_t_line_F06E02_L2_256_
D1_e20 [2018-04-23 13:00:50]
Average blind predict for 10202 obs
of 1 steps: 0.83
Average blind predict for 1336 obs of 2 steps: 0.74
Average blind predict for 139 obs of 3 steps: 0.29
Average blind predict for 18 obs of 4 steps: 0.33
Overall R2: 81.6 based on blind predict
Average train predict for 10202 obs of 1 steps: 0.83
Average train predict for 1336 obs of 2 steps: 0.15
Average train predict for 139 obs of 3 steps: 0.03
Average train predict for 18 obs of 4 steps: 0.01
Overall R2: 74.1 based on train predict
```

Conclusion

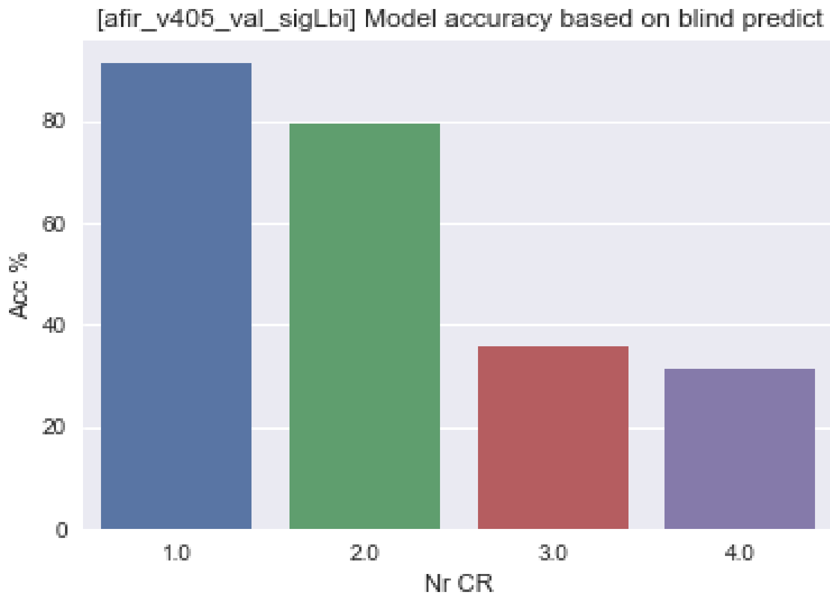
Specifically, result (A) represents the ability of the neural model to predict all steps without being helped at each step with the actual data, while result (B) represents the ability of the model to predict steps using at each new step as input the reality found in the previous step.

Example: Let's say our 20,000 EUR project is with two CRs: CR 1 10 days after signing the contract worth 15,000 EUR and CR 2 53 days after CR 1 with a value of 5,000 EUR.

A. In this variant, the system predicts with a capacity of 62% the value of 10 days and the amount of EUR 15,000, as well as the final amount of EUR 5000 and the value of 53 days between CR1 and CR2. The system stops the prediction cycle when the total sum of CRs reaches EUR 20,000.

B. In variant (B) the system will self-check at each step with the actual data of the previous step. Specifically, let's say that for step 1 he predicted 10,000 euros every 30 days. For step 2, however, we will not provide these values to the system, but the REAL values found by the operator – namely 15000 to 10 days. Thus, the system will adjust its prediction for step 2 according to the correct values.

Analyzing the results, we notice an average value of over 80% of the prediction capacity of reimbursement requests, but this tends to be very good for projects with a single request compared to projects with multiple requests due to the lack and inconsistency of data.



Source: Source:Sistemul financiar contabil al AFIR.

1. The system should take into account and analyze projects that are ongoing differently from projects that are completed. A project that is completed must be "understood" by the neural model both if the beneficiary has asked for reimbursement of all the money and if the funds have been partially settled. The introduction of both ongoing and completed projects into the neural training process introduces major confusion in neural models.
2. The amount of training data is extremely important. In the architecture phase of neural models, the dataset used total was about 11 thousand time series, of which about 90% were single-shot (CR). Cleaning up a database with as much historical data and as many cases as possible that have an acceptable amount of information will invariably lead to an ensemble of neural models that will predict multiple patterns of projects with very high accuracy.
3. In later phases, semantic models can be constructed, than can capture and present different patterns projects according to the actual behavior of projects in that category. A first prototype of this concept has been used in the present project to determine links and similarities between different types of projects within neural model optimization.
4. The entire PoC can be transformed into a functional and operational system through web services and user interface following a subsequent development cycle.
5. PoC also demonstrates high scalability and flexibility in applying for multiple measures, sub measures, appeals, etc.
6. In a potential production system, mixtures of neural graphs will be used that will have the ability to generate much better than the models tested in the PoC project.

The AI role in the financial process of the institution enhance efficiency, speed, accuracy and security.

References

AFIR data base

Bergerman M., Billingsley J., Reid J., van Henten E. (2016) Robotics in Agriculture and Forestry. In: Siciliano B., Khatib O. (eds) Springer Handbook of Robotics. Springer, Cham, Springer-Verlag Berlin Heidelberg 2016, DOI https://doi.org/10.1007/978-3-319-32552-1_56

CECCAR, (2019), Profesia contabilă într-o economie bazată pe globalizare şi digitalizare (I), https://www.ceccarbusinessmagazine.ro/profesiacontabila-intr-o-economie-bazata-pe-globalizare-sidigitalizare-i-a5481/

Ciro Potena, Daniele Nardi, Alberto Pretto - Fast and Accurate Crop and Weed Identification with Summarized Train Sets for Precision Agriculture, Springer International Publishing AG 2017, in W. Chen et al. (eds.), Intelligent Autonomous Systems 14, Advances in Intelligent Systems and Computing 531, DOI 10.1007/978-3-319-48036-7_9

Gonzalez-de-Santos, P., Ribeiro, A., Fernandez-Quintanilla, C. et al. - Fleets of robots for environmentally-safe pest control in agriculture, Precision Agric (2017), 18: 574, https://doi.org/10.1007/s11119-016-9476-3

Duţescu, A., Mihai, M.-S. (2023), The External Technological Context of Artificial Intelligence in Financial Services, Audit Financiar, vol. XXI, no. 4(172)/2023, pp. 747-759, DOI: 10.20869/AUDITF/2023/172/025

Gangu Naidu Mandala, Dharam Buddhi, Mahalakshmi Armugam, Suhas Harbola (2022), A Critical Review of Applications of Artificial Intelligence (AI) and its Powered Technologies in the Financial Industry, pp 2362-2365

https://docs.microsoft.com/en-us/azure/virtual- machines/windows/sizes-gpu

https://www.oracle.com/ro/erp/ai-financials/what-is-ai-in-finance

https://docs.microsoft.com/en-us/cognitive-toolkit/using-cntk-multigpu- support-with-keras

https://blogs.technet.microsoft.com/machinelearning/2017/10/26/anaconda-and-microsoft-partner-to-offer-python-and-r-for-powerful-machine-learning/

Larry D.Wall (2018), Some financial regulatory implications of artificial intelligence, Journal of Economics and Business (Elsevier), Vol.100, pp 55-63

McKinsey, (2020), The state of AI in 2020, https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2020

Nicolay Lomakin, Maxim Maramygin, Alexander Kataev, Sergey Kraschenko, Olga Yourova, I.Lomakin (2022), Cognitive Model Of Financial Stability of the Domestic Economy Based on Artificial Intelligence in Conditions of Uncertainty and Risk, International journal of technology, Vol 13, Iss 7, pp 1588-1598

Pazarbasioglu, C. , Garcia, A., Uttamchandani, M., Natarajan, H., Feyen, E., Saa, M., (2020). Digital Financial Services, World Bank Group, https://pubdocs.worldbank.org/en/230281588169110 691/Digital-Financial-Services.pdf

Ricardo Soto, Juan A.Gomez-Pulido, Stephane Caro, Jose M.Lanza-Gutierrez, (2019) Data Science and AI-Based Optimization in Scientific Programming, Scientific Programming (Hindawi), Vol 2019, pp 1-3

Sha Yuan, Zhou Shao, Xingxing Wei, Science behind AI: the evolution of trend, mobility, and collaboration, (2020), Scientometrics (Springer International Publishing), Vol.124, Iss:2, pp 993-1013

Appendix

This paper was co-financed by The Bucharest University of Economic Studies during the PhD program.