Computational analysis of the digital footprint using machine learning and artificial intelligence

V D Munister1,2,7, A L Zolkin3,8, V N Malikov4,9, O V Kosnikova5,10 and I A Poskryakov6,11

1 Mariupol State University, Mariupol, Donetsk region, Ukraine
2 Department of Enterprise Economics, Donetsk National Technical University, Pokrovsk, Donetsk region, Ukraine
3 Computer and Information Sciences Department, Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia
4 Department of General and Experimental Physics, Altai State University, Barnaul, Russia
5 Department of Economic Cybernetics, Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, Russia
6 Department of Labor Economics and Human Resources Management, Russian Transport University, Moscow, Russia

E-mail: tilandiya@yandex.com7, alzolkin@list.ru8, mirotnas@gmail.com9, oksana_1209@mail.ru10, i.poskryakov@miit-ief.ru11

Abstract. The article discusses the procedure of step-by-step formalization of the software model of the system for accounting and forecasting the effectiveness of employees of an IT enterprise. The issue of control of social interaction is considered. The relational model from game theory based on a specific data model template is proposed. Generalization in form of highlighting the categorical apparatus of metrics that directly or indirectly affect the procedure for assessing work efficiency, downtime or incorrect use of a working device (computer) is proposed. An architectural model of the application is proposed and substantiated, a model of a system for working with metrics is determined, the implementation of the necessary analysis tools in form of a combination of various machine learning algorithms used in systems with a binary classification based on decision making by the operator (the object of analysis) is described. The article summarizes the economic effect of the implementation of this approach in employee control systems.

1. Rationale
At the moment, there is a need of development of unified strategy (planning model), which will determine the direction of development and technical regulation of business systems, taking into account regional characteristics and modern cybernetic tools for promotion of business ideas in the corporate segment of the market from the point of social interaction and personal components [1].
Often, the level of control of the spheres of the enterprise directly related to a person is the main criterion for the high efficiency of an enterprise in the post-industrial sphere. Operational processes are reconciled and stabilized, and the role of a person, as an operator and an ultimate beneficiary, is important, deterministic and, at the same time, difficult to assess and regulate. The degree of elaboration and adaptability of man-machine interfaces for control and monitoring, adjusted ergonomics of the working space, biological commensurability play an important role in the amortization of various costs and minimization of time costs [2].

The author’s comprehensive project on accounting and dynamic forecasting of the performance of an IT enterprise is considered in the article, through the implementation of an analytical software platform for assessing the effectiveness of employees by analyzing a digital footprint and other methods, directly with data analytics during their work using computers.

2. Research cybernetic system with functions to review process information and provide a convenient human-machine interface

Thus, the subject of research is a cybernetic system with the functions of selecting information about operational processes and providing of a convenient human-machine interface for analyzing the history of the process and implementing automatic control of the process through the methods of machine learning and artificial intelligence. Before proceeding to the choice of the method of designing of an automated system, the object of computerization and structural units shall be defined in the form of the categorical apparatus of the "digital footprint" data.

In this case, the object of automation is the calculations that have been previously carried out manually by various personnel services.

The object of analysis is data traffic within the corporate network of an enterprise and other telemetric information (“digital footprint”) [3].

The analysis is based on the data of the timesheets from the personnel department, namely, the parametric characteristics of the analysis were highlighted (by the deductive method).

Thus, the categorical apparatus of data was reduced to:

1) calculation and analysis of the use of working time;
2) calculation of internal performance indicators and further ranking of personnel: workload, speed of work, complexity of work, volumetric component in work, success, and timeliness [4];
3) accounting for the range of confirmed competencies of employees to take into account the training needs and the further formation of specialized training programs [5];
4) fixing operating drivers for assessing the costs of project implementation [6];
5) assist the manager in generating feedback for employees on the results of their work and in making personnel decisions.

Requirements for an automated accounting system are reduced to the requirements of the following items:

1. availability of a control device with a web application with a simple and intuitive interface;
2. the ability to assess the effectiveness in the context of divisions, groups, personalities;
3. scalability of the system up to the "Department" level and higher.

Most often, such applications are based on the MVC / MVVP design pattern (see figure 1), which fundamentally streamlines the functional development of SQL-oriented intelligent systems.
Figure 1. MVC-architectural template of a system for accounting and forecasting the performance of enterprise employees based on the analysis of a digital footprint.

The developed service shall provide for the following roles for external users:

- technical user - an account that performs automatic authorization between external services;
- administrator - an account that allows to configure the service and all its subsystems;
- developer - the ability to develop, administer and control;
- employee - use of the system within the framework of the performance of labour and production functions;
- the head of the department - use of the system within the framework of the performance of labour and production functions;
- HR manager /financial specialist- use of the system within the framework of the execution of labour and production functions.

In case of hardware failures, including an emergency power failure, the information system shall automatically restore its operability after the failures are eliminated and the hardware restarts correctly (except in cases of damage to working media with executable program code) [7].

3. Schematic diagram of the developed model of employee performance metrics management

Based on the above, it has been decided to create an MVC data model for the UserGames session of a single employee. In other words to represent the digital footprint of the employee to the basis of social contact from the point of view of game theory [8].

The conceptual data model is reduced to five "fields" - to the basic data of the user, the game mode (work process, that is, the process of the tracking activity dynamics), the "game" itself (the process /project, interaction with the working group and other persons) (see figure 2).
Figure 2. Data model of the designed system.

The logic of the automated system: the general principle of the solution is to regularly upload data to its own relational knowledge base. Analysis and output to web pages, Excel tables, or reports sent to the mail are carried out from this knowledge base.

A stable system shall provide protection against unauthorized access due to the design basics of MVC and a secure connection (HTTPS + SSL + TLS 2.0) without reference to the standards for category 1G according to the classification of the current guidance document of the Federal Service for Technology and Export Control of Russia “Automated Systems. Protection against unauthorized access to information. Classification of automated systems” [9].

The described components of the accounting and forecasting system provide: user identification and authentication.

And as part of ensuring compliance with the requirements for the protection class, the system:

1) shall carry out identification and authentication of access subjects during entrance to the system by an identifier (code) and a conditionally permanent password with a length of at least eight characters;
2) identification of programs, volumes, directories, files, records, fields of records by names shall be carried out;
3) access control to protected resources shall be carried out in accordance with the access matrix within the information security subsystem;
4) registration of entry to/ exit from the system, or registration of loading and initialization of the operating system and its software shutdown shall be carried out. Logging out of the system or shutdown is not carried during automated system hardware shutdown.

It is allowed to expand the above mechanisms of protection against unauthorized access to achieve their compliance with the modern technological level due to BigData and machine learning.
The performance management system (see figure 3), will automatically restore its functioning after an accident with a correct hardware restart due to the implemented machine learning tools, as well as data replication and approximation systems. The system provides the ability to organize automatic or manual data backup.

Figure 3. Schematic diagram of the developed model of employee performance metrics management.

The system for data processing and prediction of the efficiency curve is based on a machine learning algorithm, which implies a prediction of the occurrence of events based on the probabilities of the onset of a peak in scalar values (in this case) using the complex method of the XGBC classifier used in binary classification problems, which is very convenient in systems mediated with the speed of decision-making (switching on, off, manufacturing, completion of the project stage) from the point of the people who are the sources of analysis.

The following objects and classifiers have been used as an analysis tool, and the standard tools of the matplotlib library have been used as a visualization tool.

#Common Model Algorithms
from sklearn import svm, tree, linear_model, neighbors, naive_bayes, ensemble, discriminant_analysis, gaussian_process
from xgboost import XGBClassifier

#Common Model Helpers
from sklearn.preprocessing import OneHotEncoder, LabelEncoder
from sklearn import feature_selection
from sklearn import model_selection
from sklearn import metrics

#Visualization
import matplotlib as mpl
import matplotlib.pyplot as plt
import matplotlib.pylab as pylab
import seaborn as sns
from pandas.tools.plotting import scatter_matrix
#Configure Visualization Defaults

```python
#%matplotlib inline = show plots in Jupyter Notebook browser
%matplotlib inline
mpl.style.use('ggplot')
sns.set_style('white')
pylab.rcParams['figure.figsize'] = 12.8
```

It shall be noted that the declaration provided above uses an alternative approach (boosting), in which each recruiter is based on the assessment of the candidate by the previous specialist. This speeds up the interview process, as unsuitable candidates are immediately weeded out. This method is used to clarify the accuracy of the data for new employees [10,11].

The developed intellectual property product (the information and analytical module) was required to have a simple and understandable web interface in the form of separate pages for each created tool.

It was decided to perform it in the form of some visualization of a SCADA-like system.

Each created tool shall be able to provide the view of information from the department level to the level of an individual employee in the form of graphs [12].

The developed automated system shall have at least three modes: regular PC version, mobile version and version for wall monitors [13,14,15].

Also, the system itself shall be developed taking into account the provision of its further development and functionality increasing. At the same time, the main architectural principles of the system shall already be laid down in it, allowing for its further development, primarily associated with the addition of interaction with other products and systems. In the future, the system shall allow expanding to other divisions and branches.

4. Findings

With the use of intelligent accounting system, the efficiency of employees increased by 20%. The total number of employees of the IT enterprise is 300 people, the average salary, including contributions is 2 million rubles per year, thus, the economic effect is 0.2 * 300 * 2,000,000 = 120,000,000 rubles.

Thus, the use of modern architectural design patterns of web applications in combination with machine learning methods in the field of data analysis (metrics) signalling the level of employee performance gives a good economic effect and, in general, leads to step by step minimization of unnecessary downtime and ineffective steps by screening various "digital" movements in the analysis of corporate network activities.

References

[1] Dolezilek D et al. 2003 Secure scada and engineering access Communications: a case study of private and Public communication link security *Schweitzer Engineering Laboratories Inc. Pullman, Washington USA*

[2] Koushik A N and Rashmi B S 2016 4 th Generation SCADA Implementation for Automation *International Journal of Advanced Research in Computer and Communication Engineering* 5(3)

[3] Tan S Y and Taeihagh A 2020 Smart City Governance in Developing Countries: A Systematic Literature Review *Sustainability* 12(3) 899 doi:10.3390/su12030899 URL: https://ssrn.com/abstract=3526607

[4] Tonkih E V, Paramonov A I and Kucheryaviy A E 2020 Analysis of wireless high-density IOT network *Electrosyvaz* 1 (Moscow: Info-Electrosyvaz LLC) 44-8

[5] Hu H and Li Y 2020 Remote status monitoring based on ESP8266 connected IoT cloud platform *IOP Conference Series: Earth and Environmental Science – IOP Publishing* 461(1) 012022

[6] Kurundkar S et al. 2017 Remote monitoring of solar inverter (an application of IOT) *Am. J. Eng. Res.(AJER)* 6(7) 70-4
[7] Tadeus D Y et al. Prototype of wireless sensor network based on open industrial protocol: modbus. IOP Conference Series: Materials Science and Engineering – IOP Publishing 801(1) 012140

[8] Alves T and Morris T 2018 OpenPLC: An IEC 61,131–3 compliant open source industrial controller for cyber security research Computers & Security 78 364-79

[9] Jagadeesh M et al. 2019 Monitoring system in industry using IoT 5th International Conference on Advanced Computing & Communication Systems (ICACCS) – IEEE 745-8

[10] Kumar M et al. 2021 XGBoost: 2D-Object Recognition Using Shape Descriptors and Extreme Gradient Boosting Classifier Computational Methods and Data Engineering (Springer, Singapore) 207-22

[11] Akhmetshin E M, Kovalenko K E, Mueller J E, Khakimov A K, Yumashev A V and Khairullina A D 2018 Freelancing as a type of entrepreneurship: Advantages, disadvantages and development prospects Journal of Entrepreneurship Education 21(Special Issue 2)

[12] Tormozov V S, Zolkin A L and Vasilenko K A 2020 Optimization of neural network parameters based on a genetic algorithm for prediction of time series International Multi-Conference on Industrial Engineering and Modern Technologies FarEastCon 2020 9271536 DOI: 10.1109/FarEastCon50210.2020.9271536

[13] Lavrov E A, Zolkin A L, Aygumov T G, Chistyakov M S and Akhmetov I V 2021 Analysis of information security issues in corporate computer networks IOP Conference Series: Materials Science and Engineering (Krasnoyarsk Science and Technology City Hall, Krasnoyarsk, Russian Federation) 12117 DOI: 10.1088/1757-899X/1047/1/012117

[14] Vasilev V L, Gapsalamov A R, Akhmetshin E M, Bochkareva T N, Yumashev A V and Anisimova T I 2020 Digitalization peculiarities of organizations: A case study Entrep Sustain 7 3173-90

[15] Zhang T, Shaikh Z A, Yumashev A V and Chłąd M 2020 Applied Model of E-Learning in the Framework of Education for Sustainable Development Sustainability 12 6420