SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2023
ORGANIZATION AND MANAGEMENT SERIES NO. 177

BUSINESS ANALYST – A POSITION ATTRACTIVENESS AND MARKET REQUIREMENTS, A SAMPLE FROM POLAND

Mieczysław PAWŁOWSKI[1*], Dorota WALENTEK[2]

[1] Maria Curie-Sklodowska University in Lublin, Faculty of Economics; mieczyslaw.pawlowski@mail.umcs.pl, ORCID: 0000-0002-9693-0573
[2] Czestochowa University of Technology, Faculty of Management; dorota.walentek@pcz.pl, ORCID: 0000-0003-2943-8359
* Correspondence author

Purpose: The aim of the study is to classify job offers for analyst positions according job title and job requirements and, on this basis, to use machine learning methods of text mining to build job profiles for selected job categories.
Design/methodology/approach: Expert classification and machine learning algorithms.
Findings: The study shown specific set of keywords for business and financial analysts, financial controller, data scientist, system and security analysts. Research also discovered the coherence relationship of job title and job responsibilities and how job title is used for prestige.
Research limitations/implications: Research is limited to sample of data.
Practical implications: The study deliver multidimensional knowledge for managers.
Social implications: The study describe an analyst position as the sets of skills expressed by keywords completed based on job level, job title and job requirements and by that helps different groups of people as for example students to allocate their attention properly to the market and the industry they match.
Originality/value: We use ensemble strength of expert judge and machine learning. Using expert approach we recognized job categories base on text summary prepared by ML and then continue job offer classification by applying algorithmic text mining.

Keywords: business analyst; text classification; job position, machine learning.

Category of the paper: Research paper.

1. Introduction

Business analysis is increasingly used in organizations, especially those whose activities are based on the processing of various types of operational data in transactional systems (Goh et al., 2019). J. Park and S. Jeong (2016) recognize business analytics as an element necessary for the success of a project related to IT systems. According to the Global State of Business

Analysis Report, created by the International Institute of Business Analysis, as many as 46% of organizations have increased their involvement in projects based on data analysis (IIBA, 2021). Canges in organizational structures caused by the COVID-19 pandemic and the global increase in the level of cyber threats have contributed to the increased demand for specialists in the field of data analysis, agile analysis and digital initiatives.

As of March 1, 2023, the Web of Science database contained a total of 4,696 publications on this subject of business analytics, and the Scopus database contained 600 publications. However, there are much fewer publications devoted to the business analyst (WoS: 61, Scopus: 449). For years, the theoretical study A Guide to the Business Analysis Body of Knowledge, created by the International Institute of Business Analysis (IIBA, 2015), was considered a compendium in the discussed area. This study was the basis for research conducted by J. Park and S. Jeong (2016). However, their literature review and conducted research were limited to the years 1990-2016. They can be treated as an important basis for further research, but the results of research published in 2016 should not be considered up-to-date from the perspective of people managing the organization in 2023. studies on the role of a business analyst in an organization are most often theoretical in nature (Gobov et al., 2020).

Pracuj.pl reports that the pandemic break shown a little stagnation in recruitment activity in the finance and banking sector, but now they become again the most frequently recruiting industries. In 2022 finance sector overtook the IT industry of 21%, where IT was the leader of the year 2021. The jobs offered by financial sector counts 12% (130,000) of total offers published in Pracuj.pl where IT published 107,000 offers. IT and finance together shown the dynamics of 103% y/y. Pracuj.pl reports that the demand for different digital based and data mining specialists are increasingly growing among wide range of sectors, even for which the area of technology is not the main profile of business activity. That is why, we were motivated to investigate the job offers characteristics where the demand is placed for analytical skills in IT and finance sectors and then convert it to the model for that kind job's skill description for other industries.

The aim of the study is to classify job offers for analyst positions according job title and job requirements and, on this basis, to use machine learning methods of text mining to build job profiles for selected job categories. An attempt was made to answer the following questions:

1. What the analyst position's categories can be extracted from the titles of employment offers. What are the most important phrases offers have used? How we can define "the most important phrases" Are they the differentiation factors between job categories?
2. What is the persona profile in selected categories of business analyst. What kind of skills should the offeror have? What responsibility employers allocated to those positions and what benefits they offers?

We decided to select IT and finance as the mature sectors, which can deliver a most solid and reliable job description on the certain position. We were also interested if we can investigate that kind of dataset by utilization text mining methods. Thus, the whole study brings clear benefits for academia and practitioners. As researchers, we can better understand how the analyst-kind job is defined, described and latently understand by industries. The industry practitioners can utilize this knowledge to understand more precisely the area of skills other company tries to cover and then better express their demand for candidate profile. In addition, it should be emphasized that among numerous scientific publications devoted to the subject of business analyst, there is definitely a lack of publications devoted to the requirements of employers with regard to the position of a business analyst in the current market reality. Our research presents the current requirements of employers, in line with market realities.

2. Methodology

2.1. Flowchart

The goal of our research is to create the profile of business analyst as the job position now and its projection for the future. Dataset has been built based on collection of active job offers published in the most popular polish recruitment portal "Pracuj.pl". According to the their report they have published 1,103,000 offers in 2022. It makes them an ultimate market leader in Poland. In the report they pointed that there is clearly visible demand of digital competences in the candidate profile. This includes not only pure IT and financial competences but also analytical, data mining and data science as becoming a core competences in numerous sectors. This trend can be noticed not only in IT and Finance but in the range of sectors. Another observation they signaled was a tendency in work organization. A full remote or hybrid work organization becomes a constant proposal for new employment. This change can indicate a progressing growth in the flexibility of work organization in companies where the job done is more important than physical presence in the office.

Because there are some variation of business analytics position understanding in different industries we must detect those differences and then create the analyst persona for each category. The whole analytical flow and data research is presented in figure 1.

The analysis is based on dataset composed from the internet job offers, collected by using a web scrapping. All data a publicly available from the internet browser without any brake of the login or other authentications or security features.

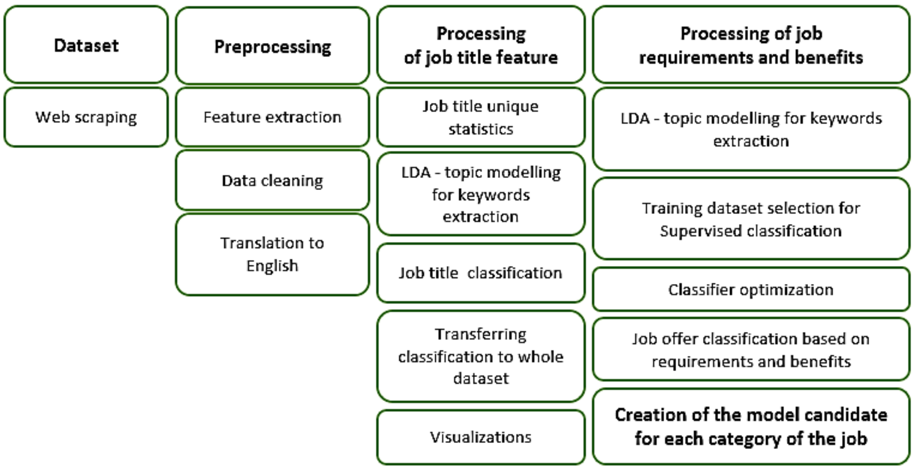

Figure 1. The flow of data research.

Source: own work.

In our research, we have collected job offers from categories IT and Finance with additional filtering for phrase “analyst”. The offers can be distinguish between two groups. The first group, includes the majority of the offers, which is published in form of short description in the listing page and comprehensive description in the offer’s card. The second group of nearly 20% of total offers in the sample included the offers published only as the reference to the offeror’s web page where the offer was encapsulate inside company website. Because of individual composition of every offer they have been excluded from dataset. Finally we have collected a sample of 3,467 unique offers from the period of 6 weeks. The final sample for the research represents 1.3% of total offers in categories IT and Finance. Because we cannot calculate our sample relation to total population of offers the result of our investigation cannot be interpreted as representation of phenomenon.

In the analytical phase we have made data cleaning, feature extraction and translation to English using Google translate. Finally we have got a dataset consistent of following columns: pos_href, pos_title, pos_level, job_offer_cart, pos_level_en, pos_title_en, responsibilities, requirements, offered, technologies, training-space, and benefits. We believe the columns meaning is understandable without further explanation.

First do a selection of unique job titles and position levels for clustering. The goal of the clustering procedure was to classify job titles based on keywords we have extracted from job titles. We have utilized Latent Dirichlet Allocation (LDA) procedure (Blei et al., 2003) from Java based implementation Mallet (Shawn et al., 2012). Since we have job offers categorized we have transferred them to all records in dataset. We documented all operations and changes occurred in data in tables and visualizations.

The next analytical phase was to study the requirements and benefits of the job offer and check how they feet to category build based in previous phase. We proposed this direction of the research because job title is straightforward job description, limited in size and that is why relatively easier to classify. The risk is if the job requirements are following job title, therefore we processed also additional features from the dataset for confirmation or correct previously setup classification.

In the third step we build classification model for job requirements and job benefits, optimized classifier and finalize classification taking to account: job title, requirements and benefits.

Finally, we build job personal for all categories, what was the idea and goal of the research.

2.2. Classifier optimization

For classifier optimization we have used FLAML, which is a Python library (https://github.com/microsoft/FLAML) designed by Microsoft to optimize machine learning models with low computational cost. It supports machine learning and AI tasks like NLP classification, which we were used in our study (Wang et al., 2021; Liu, X., Wang, C. 2021).

3. Results

3.1. Position level

We have analyzed demand for the particular position levels. A major demand is visible for the regular specialist. The second group is senior and junior specialist. They appears in almost equal frequency of 17-18% each (35% together) in job offers. Therefore, the top-three box score is equals 91% of total. The remining 9% of demand is distributed among the managers, trainee, and assistants. For details see table 1.

Table 1.
Position level statistics

Position level	Count	%
specialist (mid regular)	2,361	57%
senior specialist (senior)	740	18%
junior specialist (junior)	687	17%
expert	169	4%
manager coordinator, director	113	3%
apprentice trainee	48	1%
assistant	23	1%
	4,141*	100%

Note *) the quantity 4,141 excide total number of records in the sample (3,467) because a some number of offers include a double options for position level.

Source: own work.

The count 4,141 excide total number of records in the sample (3,467) because some number of offers includes a multiple options for position level.

3.2. Job title statistics

As mentioned, before processing the analysis we have made dataset preparation in order to eliminate punctuation, stopwords and special characters. We also removed double wording like male and female names for the position which in English is the same word. As the next step we have applied NLTK's WordNetLemmatizer for lemmatization. Finally, we have built the job title for the analysis from the list of lemmas. Total unique job titles equals 1,837.

The jobs titles statistics is presented in figure 2. There are only 8 job offers where the job title consistent of only one word, like "analyst" or "accountant", 224 offers for 2 word jobs like "business analyst" or "data scientist", 630 offers includes 3-words job title like "business intelligence specialist" or "financial controlling specialist", and 512, 4-words job's titles like "power bi data analyst" or "business analyst sale support". As the conclusion one can note that the majority of job titles consists of 3 or 4 words. The details can be seen from figure 2.

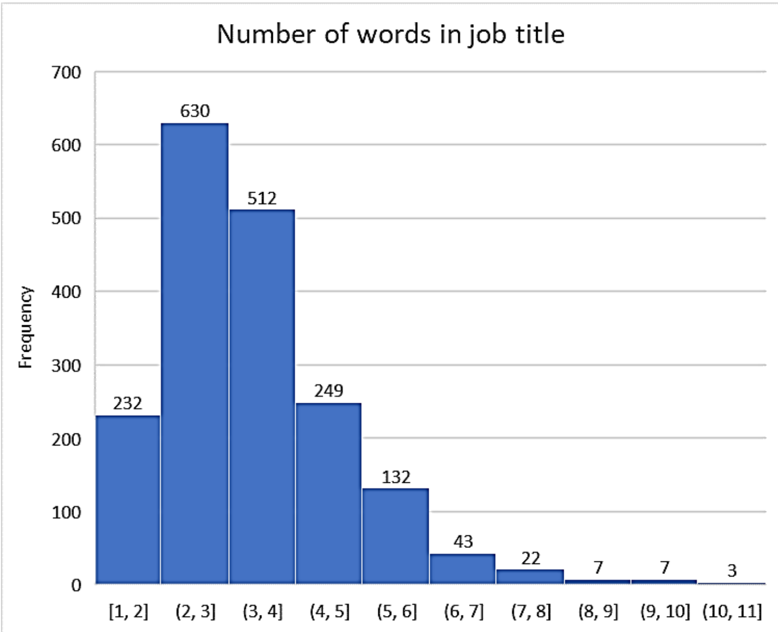

Figure 2. Number of words in job title.

Source: own work.

The statistics of the job positions demand in the sample of job offers shows that the top 10 position in a list is occupied as follows: "financial analyst", "business analyst", "financial controller", "accountant" , "controlling specialist", "data analyst", "data engineer", "business system analyst", "data scientist" and "analyst". As we can see, the top demand goes to financial and business analyst but also accountant and financial controller and controller specialist have go numerous demand, however we have been searching for "analyst". Perhaps some comment con be placed to "accountant" which is seen partially as an analytical position.

3.3. Bigram analysis

Bigram analysis shows words collocation. We have built bigrams based on job position titles. The top 20 bigrams is presented in figure 3. One can see that we have clear collocations between an “analyst” and “business”, “system”, “financial” and “controller”, but also “finance” and “accounting”. An “analyst” is also linked to “risk”, “credit”, “it” and “support”. Then we have node “data” in between and collocation to “scientist” and “engineer”. Through node “analysis” we have got a “specialist” of “settlement”, “controlling” and “reporting”. A collocation analysis of “annalist” with its specialization is presented in figure 4. One can see that “analyst” has a strong collocation to the nodes mentioned before but also to “sales”, “project”, “planning”, “reporting”, “operations”, “client”, “management” and “market”. Therefore we can conclude that analyst position is expected as to person with the skills extended to different area of business operations and the position is situated in different departments.

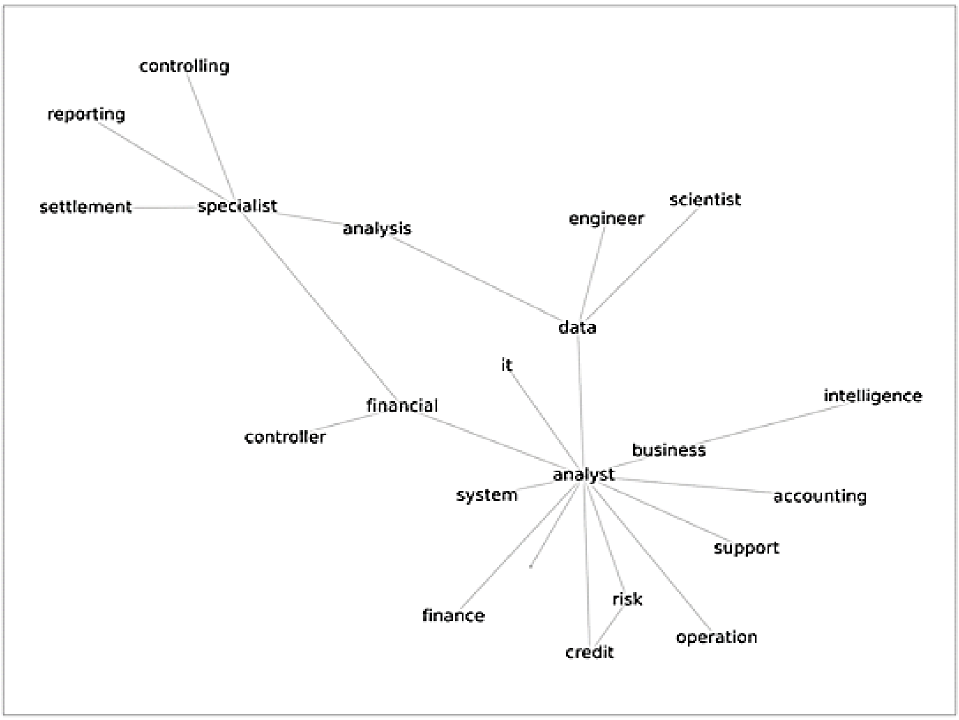

Figure 3. Top 20 bigrams.

Source: own work.

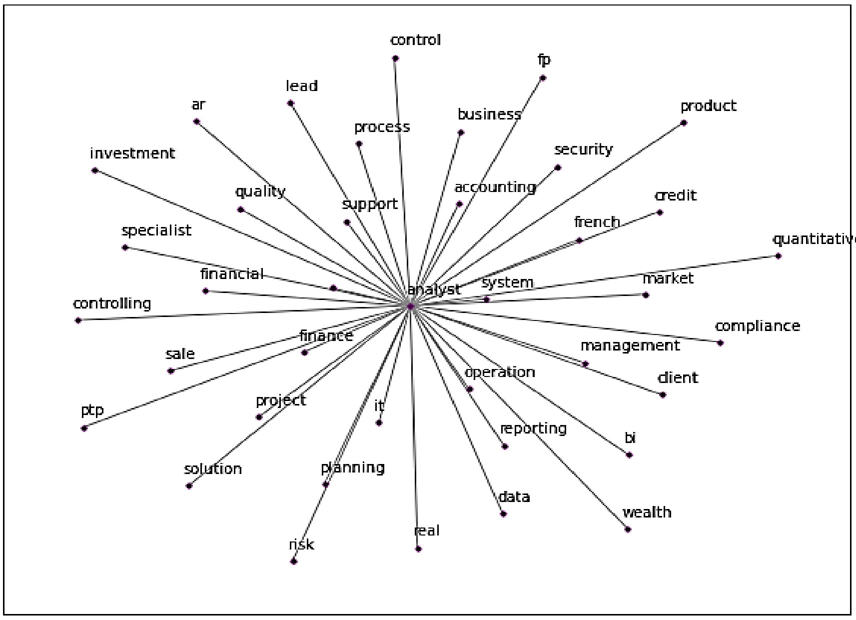

Figure 4. Direct collocation of "analyst" with to 50 specialization.

Source: own work.

That means that people employed on the position of analyst should investigate very deeply the specifics of the certain area of the business where they are working. This make challenge not only for education of such students but also for themselves, as employee for self-education on on duty on the job position or advanced onboarding. This, in turn, require a high self-organizing skills, willing and ability to extend education in the moment they it just finished. This also means for the company, to have an organizational culture with high tolerance for errors during self-education or huge education budget. It is easier make onboarding if the company has an analytical team, but if it that is the new employed person on the empty ground, the future depends of survival skills.

3.4. The wording structure in the job tittle

Based on Mallet topic modeling (Shawn G, et al. 2012), we have selected eight main category of analysts: "business analyst", "data scientist", "financial analyst", "financial controller", "intern analyst", "security analyst", "system analyst" and "other". We prepared other category if the process cannot classify a job to any of named category.

Mallet is a Java-based application for language processing, including document classification, clustering, and topic modeling. It uses different variants of Latent Dirichlet Allocation for topic modelling. Mallet process lest for topic modelling in two steps: the topic importing from the raw data set and the topic training. The only thing we must tell the Mallet is the number of topics we want to test. The rest parameters like number of iterations or smoothing are adjusted during internal optimization. Finally we have got the results contains the topic key words, and the matrix of the most prominent documents for each topic.

We have tried different number of topic from 5 to 60. If there is too little topics selected then we got not clear theme and consistency visible form keywords in each topic. I case where we have got to may topics the wording is repeated with a little variations. The most clear keywords, consistent internally which looked logically dispersed we have got for 50 topics. The several topics for illustration is presented in Table 2. but most important wording in each category in form of word cloud is presented in figure 5.

Table 2.
Illustration for topic modelling

Topic
business analyst - topic 0 * analyst (232.0) market (121.0) data (101.0) power (71.0) trading (61.0) energy (51.0) cee (30.0) sustainability (20.0) researcher (20.0) trainee (20.0)
business analyst - topic 1 * process (187.0) automation (140.0) robotics (62.0) trainer (62.0) expert (47.0) rpa (47.0) operational (47.0) tool (47.0) tester (31.0) robot (31.0)
financial analyst - topic 2 * analyst (235.0) credit (163.0) risk (146.0) management (72.0) team (42.0) expert (40.0) model (30.0) collection (20.0) retail (20.0) language (17.0)
data scientist - topic 3 * data (294.0) scientist (163.0) engineer (101.0) lead (70.0) learning (39.0) machine (31.0) pricing (16.0) banking (16.0) workday (16.0) delivery (16.0)
system analyst - topic 4 * sap (137.0) center (59.0) key (59.0) user (59.0) commerce (39.0) planner (39.0) field (39.0) inhouse (39.0) competency (39.0) hana (39.0)
financial analyst - topic 5 * excel (67.0) resilience (67.0) national (67.0) hire (67.0) financial (1.0) analyst (1.0) accountant (1.0) business (1.0) controller (1.0) controlling (1.0)

Note: Category is set based on wording analysis. The number in brackets indicate the "power" of word in documents cooperating to the topic.

Figure 5. WordCloud presentation for the most important key words for job category "business analyst", "data scientist", "financial analyst", "financial controller", "intern analyst", "security analyst", "system analyst".

Source: own work.

One can see the difference if formation of the job title for every of the job category. We believe that visual outlook clearly differentiate above mentioned categories.

3.5. Job categorization

We have built job category based on keywords of topic modelling. We applied expert methodology combined with machine learning tests for consistency. It was working this way, that we after keywords selection we run classification model to check selection consistency based of confusion matrix. We tried to maximize the distinctions between categories based on unique and distinct sets of keywords. Finally we have completed the most important, 128 keywords in seven categories which are presented in table 3. The database of keywords was used to job classification based on job title and responsibility area.

Table 3.
Keywords for each category

Category	Keywords	No
Business analyst	analyst, specialist, business, product, management, monitoring, market, process, service, controlling, automation, pricing, transaction, personal, project, corporate, data, center, planning, sale, risk, supply, operation, customer, system, transfer, excellence, contract	37
Data scientist	data, specialist, engineer, developer, analytics, program, analysis, scientist, sql, graduate, innovation, analytical, associate, report, reporting, lead, etl, team, forecast, cloud,	21
Financial analyst	analyst, specialist, account, accountant, settlement, accounting, tax, support, credit, financial, risk, fund, derivative, control, reporting, billing, cost, receivable, insurance, asset, class, investment, finance, management, ptp, valuation, excel, resilience	35
Financial controller	financial, controller, accountant, general, controlling, analyst, ledger, accounting, finance	9
Other	internship, audit, intern, crm, processing, consultant	4
Security analyst	analyst, security, client, aml, prevention, anti, revenue, know, soc, designer, fraud	10
System analyst	network, administrator, performance, sap, analyst, tester, computer, center, key user	12

Source: own work.

One can note, that some keywords are common for two or more profiles. For example, “analyst” and “specialist” are present in different categories, what means that they do not differentiate the job profiles. We can find “analyst” in all categories and “specialist” in “business analyst”, “data scientist”, and “financial analyst”. There is also some common wording in “financial analyst” and “financial controller”.

We have distinguished “financial analyst” and “financial controller” because we found the phrase “financial controller” explicate placed in job title in many where in other job offers called the position as “financial analyst”. We assumed that it covers some prestige of financial controller. A similar situation we can also note in “data scientist” and “business analyst”, which include common words to all kind of “analytics” but the only job name is distinctive. The common keywords are as follows: “analysis”, “data”, “reporting”, “team”, etc. We decided not to eliminate common wordings because it is quite clear for everyone that the job positions can overlap in some areas. In case of analytic position is even more natural then for others. The preliminary profiles created for “business analyst” and “data scientist” based on job titles is presented in figure 6. As we can see, there are not overlapping keywords in the top 10 keywords presented on the pictures, but the difference in the intensity of wording is clear visible.

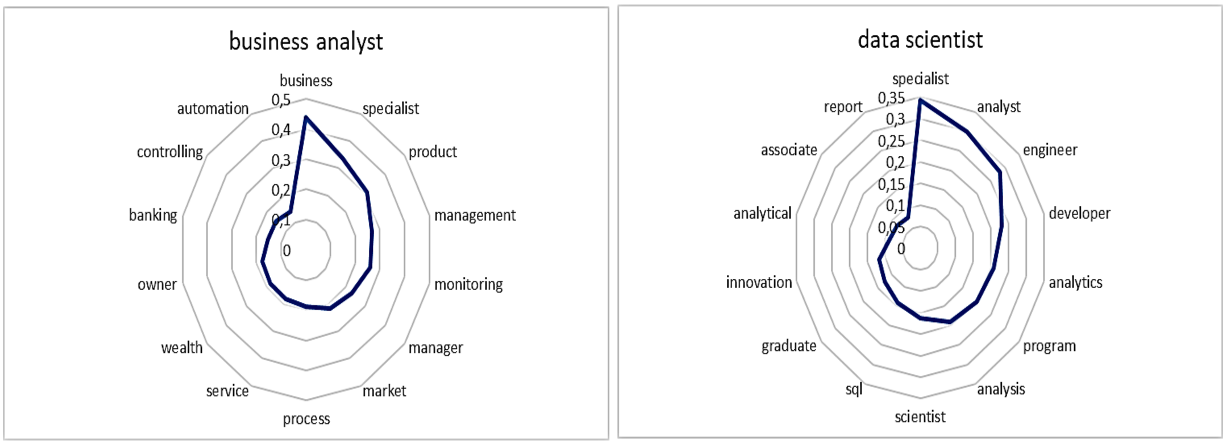

Figure 6. Preliminary profiles built base on job title word statistics.

Source: own work.

Finally, we have created job categories based on topics and then mapped job offers to categories (Table 4). We utilized mixed models of expert's and machine learning algorithms for classification. Based on that we can note that 31% of jobs belongs to category "business analyst". The next is "data scientist" – 16% and "financial analyst" – 27%. The position of "financial controller" covers 8% of jobs. The rest 2% and 7% of job offers belong to the category "security" and "system" analyst. We have got 10% of uncategorized jobs. The statistics suggests that the first four jobs are really demanding on the market and creates profile of the future of analytical sector.

Table 4.

Job classification

Category	Job offers count	% in total
Business analyst	1084	31%
Data scientist	546	16%
Financial analyst	948	27%
Financial controller	264	8%
Analyst (other)	323	10%
Security analyst	68	2%
System analyst	235	7%

Source: own work.

In next chapter we analyze what are the job description and the requirements expected by enterprises from the certain job position and how the requirements are consistent with the job title.

3.6. Job requirements classification

Job description (requirements) as a set of keywords is much longer then job title. As one can see the most frequently appearing job requirements is between 20 and 58 words, what can be seen from figure 7.

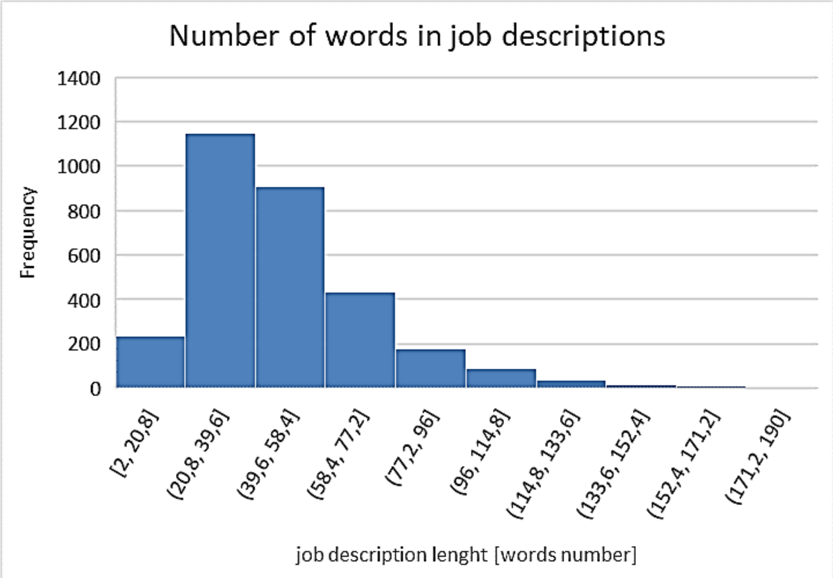

Figure 7. Job description wording histogram.

Source: own work.

Job description's bigrams cooccurrences form complicated network figure. It suggests that offers expect the responsibility for following areas: tools for controlling and analytical support, active participation in multiple projects and business processes, responsibility for business intelligence, budgeting and forecasting, decision making support, processes continuous improvement, and periodical reporting annually, quarterly, and monthly. We also can note some nodes with required responsibilities for new solutions and implementation of the projects, projects development and cooperations in internal teams or with external partners. More details you can see from figure 8, which is just the only the crop of the center part of the full picture.

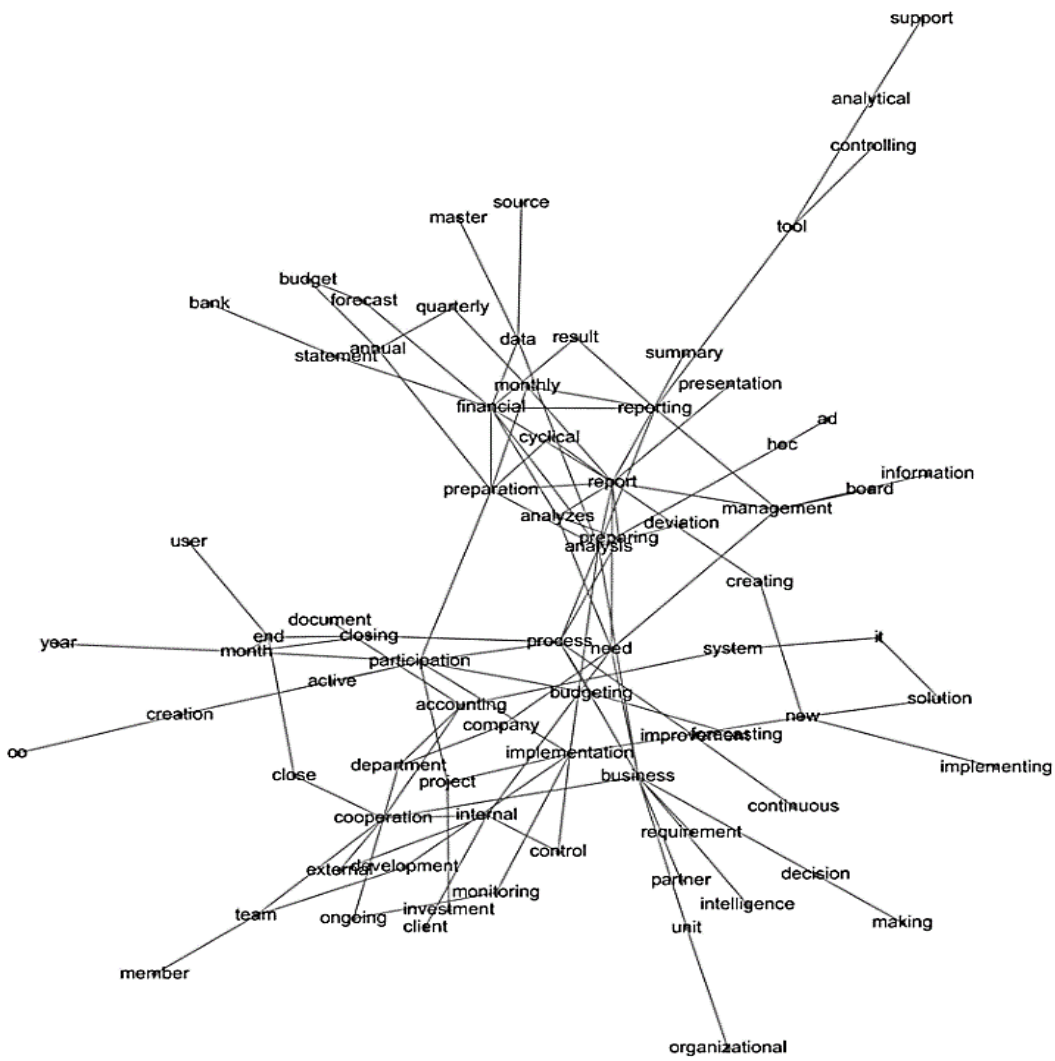

Figure 8. Job position responsibility, a central nodes.

Source: own work.

The important aspect of the job title analysis was the corresponding job description and its coincidence. If the job title is formulated in the way corresponding coherently with job description (requirements) then the reader which is a potential employee can standardize the job understanding in his imagination and relay on job title for next offer supposing that they expects skills and fields of responsibility as usual. Naturally, description of certain field of responsibility can vary from one to another industry but in the same group of firms like finance or IT the job title is expected to follow by more or less the same requirements.

As we have seen in previous chapter, job offers was allocated to seven categories based on job title topic modeling and experts work. Now we are going to check how the same categories classify job descriptions. This operation will verify profiles created based on job title.

As before, we used the similarity calculation between sets of keywords representing certain category with the set of keywords comprising job description. We used the Jaccard Index. Similarity in terms of this measure is the distance between two vectors where the vector dimensions represent the features, in our case keywords of two objects: category and job description. In our case the comparison is equal the common keywords belonging to both vectors. The biggest number of common keywords determine the category.

3.7. Job description categories

As a result of Jaccard measure we have got a matrix of job descriptions divided to categories. Now, they can be compared with the job title categorization it see how many offers can be judged as the job title is consistent with the job description. As consistent we understand that they operate the same wording belonging to the same category. Detail of comparison you can see from table 5.

Table 5.
Comparison of job title and job description categorization

Job description / Job title	business analyst	data scientist	financial analyst	financial controller	security analyst	system analyst
business analyst	862	94	114	1		12
data scientist	324	185	26			10
financial analyst	469	48	422	2		4
financial controller	138	20	105	1		
security analyst	52	7	6		3	
system analyst	195	11	9			19

Source: own work.

We clearly see that 862 offers for business analyst is consistent in title and job description. Similarly, 185 data scientists and 422 financial analysis. In this cases job searcher will experiment coherent picture of job offer. In some cases job offer description belong to other category then job title. In our case 324 job offers from title category “data scientist” belong to category “business analyst” and 26 to financial analyst when we considering job description. We didn’t find “financial controller” like wording in job description, but we found “business analyst” and “financial analyst” instead.

3.8. Job requirements

As job requirements offers describe at least experience, degree, accounting or business administration skills, analytical and organizational ability, advanced options for excel or other skills to work effectively, team environment attention, accuracy and confidence for information produced, fluency in English written spoken or other needs for a job. As example we can analyze following requirements: “bachelor degree, finance related minimum two year experience, controlling project comparable industry advantage solid knowledge, sap fi module, office tool, especially excel strong analytical skills, problem solving approach, high level interpersonal communication English” or “master degree, finance accounting or business administration minimum 3-year experience, controlling good knowledge, sap advanced, excel skills, fluency in English, written and spoken, soft analytical, paying attention, detail communication, diligence high sense, ownership ability, work effectively, team environment, time management, prioritize task well, pressure prerequisite, meeting deadline, highest quality, propose improvement process”.

As technical requirement job offerors enumerate: sql, python, jira confluence, sap, uml, bpmn power bi, enterprise architect, agile scrum, ms office, power query, active directory, panda, scikit learn, keras, pytorch, tensorflow, aws,l git, docker, apache spark, mlflow, nosql, google analytics, tag manager, azure, google cloud. More you can see in Appendix 2.

4. Discussion

The results of the study are in line with the results of research conducted by Verma (2019) and IIBA (2015; 2019): depending on the perspective and thematic scope of the person specializing in business analysis, different job descriptions are used. Common names include business analyst, business intelligence analyst, data analyst and data scientist (Verma et al., 2019). A person focusing mainly on data analysis may, in turn, be assigned the following roles: subject matter experts (SMEs), data architect, data engineer, data scientist, data analyst, data journalist and also business analyst.

Among the job advertisements we analyzed, the most common job titles were: business analyst, data scientist, financial analyst, financial controller, system analyst and security analyst. After defining the specific categories of job positions, we conclude that 31% of jobs belongs to category “business analyst”, 16% to “data scientist”, 27% to “financial analyst” and 8% to “financial controller”. The next 2% and 7% of job offers belong to the category “security” and “system” analyst. There are 10% of uncategorized jobs.

According to IIBA (2015), business analysis is a practice that enables making changes in an organization by identifying its needs and recommending solutions that are beneficial from the point of view of stakeholders. It can be strategic, tactical and operational. Business analysis does not refer to one thematic context, but to a number of different business areas and perspectives.

A related term is business data analytics. It is defined as a specific set of techniques, competencies and practices used to continuously explore, study and visualize business data (IIBA, 2019). Business analytics therefore focuses primarily on data. The purpose of this process is to gain insight into the business, which enables better evidence-based decision making. C. Aasheim et al. (2015) define business analytics as a subcategory of data analytics that is applied in a business environment to analyze related problems.

In turn, a business analyst (BA) can be defined as any person performing business analysis tasks (IIBA, 2015). The position and organizational role of this person is not important (Gosh, 2015). BA is tasked with creating and analyzing information received from, among others, from managers, other departments or units cooperating with the organization. Importantly, the business analyst is responsible for identifying the real needs of stakeholders. A frequent goal of BA is to facilitate communication between organizational units of the company by defining the scope of possibilities to meet the needs of stakeholders using ICT solutions (Chernysheva, Shepelenko, 2018). In other words, BA defines the scope of analytical work and then uses the results to support the business decision-making process and the implementation of decisions resulting from this process (IIBA, 2019). The business analyst is the key link between the customer with specific requirements and the development team offering a given solution (Shah, 2017). He is responsible not only for identifying the business problem or the organization's needs, but also for presenting these needs in a way that is understandable to the teams providing the solution (BAE, 2014).

This is consistent with the results of our research. The offers we analyzed require the responsibility mainly in the following areas: tools for controlling and analytical support, active participation in multiple projects and business processes, responsibility for business intelligence, budgeting and forecasting, decision making support, processes continuous improvement, and periodical reporting annually, quarterly, and monthly.

The BABOK study presents the Business Analysis Core Concept Model™ (BACCM™), enabling the core of business analysis, regardless of the industry, methodology, perspective or level in the organization. BACCM consists of six elements: Change, Need, Solution, Stakeholder, Value and Context. According to this model, the answers to questions about these elements are crucial for the work of a business analyst (IIBA, 2015):

- What changes are being made?
- What needs should be met?
- What solutions are being changed or created?
- Who are the stakeholders involved?
- What is the context, i.e. the circumstances that influence the change?

When analyzing the previous research on the business analyst, the focus was on publications in the Web of Science and Scopus databases as of March 1, 2023. The selection criteria were Author Keyword='business analyst', publication years: 2014-2023 and publication type: open access. The selection process for the above-mentioned publications is presented in Table 6.

Table 6.
Selection of articles using the SPL method

Search stages	**Number of publications**	
	WoS	**Scopus**
Keyword='business analyst'	61	449
Year of publication: 2014-2023	44	221
Type: open access	13	36
SUM: WoS + Scopus	49	
Removing duplicates = **final number of publications**	**43**	

Source: own work based on WoS and Scopus.

In the first step, publications with Author Keyword='business analyst' were searched. As can be seen, significantly more articles meeting the criteria were found in the Scopus database (449 in Scopus vs. 61 in WoS). Then, the search criteria was narrowed down to years of publication and its type. The total number of publications after removing duplicate items was 43. Due to such a low number of search results, the stage of title and abstract analysis was abandoned and the final content analysis was started. There were also no further narrowings, e.g. to the Management category. If, as a result of the content analysis, it was found that a given publication did not meet the search results, it was not described in this article.

Due to the fact that business analysis can be conducted from different perspectives (IIBA, 2015), the position of a business analyst was described by researchers in various areas and thematic contexts. A. Babar et al. (2018) focused on business analyst competencies. Their research confirmed that BA competencies play an important role in the requirements elicitation process. Requirements engineering in terms of business analysis was also the subject of research by M. Daneva et al. (2019). In turn, T. Chakabuda et al. (2014) assessed the competency gaps of people employed as business process analysts using the competency framework created by Sontaya and Seymour. They found that the biggest gaps concerned knowledge about the organization. Researchers also described the use of user stories by business analysts (Chauhan, 2015), the role of BA in agile ASD development teams (Ndlela, Tanner, 2022), the educational requirements for the BA position (Richards, Marrone, 2014) and the relationship between key skills for BA and the learning outcomes of students in selected fields of study (Zych, 2020). The BABOK guide (IIBA, 2015) presented BA's work from the following perspectives: agile, business intelligence, information technology, business architecture, and business process management.

From the point of view of managerial practice, it is particularly important to define the competencies relevant to the position of a business analyst. BABOK (IIBA, 2015) presents six main categories of competences: (1) analytical thinking and problem solving, (2) behavioral characteristics, (3) business knowledge, (4) Communication Skills, (5) Interaction Skills and (6) Tools and Technology. Based on these categories, J. Park and S. Jeong (2016) proposed their own version of the key competence areas for the role of a business analyst (Table 7).

Table 7.
Business analyst competencies

Area	Competencies in the mentioned area
Attitude	accountability, adaptability, ethics, time management, trustworthiness
Knowledge	business acumen, domain knowledge, methodology knowledge, organization knowledge, solution knowledge, technical knowledge
Analysis	root cause analysis, structured analysis, decision making, statistical analysis
Thinking	client experience thinking, conceptual thinking, creative thinking, learning, system thinking
Communication	Listening, non-verbal communication, verbal communication, written communication
Interaction	conflict resolution, facilitation, leadership, 481uestioning/interviewing, relationship building, teaching

Source: Park, Jeonga, 2016.

The main areas of competence indicated in the table above are attitude, knowledge, analysis, thinking, communication and interaction. Selected competencies have been assigned to each area. The authors of the study examined how important individual competencies are for the position of a business analyst. The order in the ranking is as follows (starting from the most important competence): root cause analysis, listening, accountability, leadership, ethics, written, conflict resolution, system thinking, trustworthiness, verbal communication, domain knowledge, decision making, time management, relationship building, facilitation , business acumen, questioning, conceptual thinking, organization knowledge, client experience thinking, teaching, structured analysis, statistical analysis, creative thinking, adaptability, solution knowledge, learning, non-verbal communication, methodology knowledge, technical knowledge (Park, Jeonga, 2016, pp. 3996-3997).

The competences of persons dealing with data analysis in relation to business data analytics depend on the scope of duties in a given position. The most frequently mentioned competences are: specific knowledge of the business sector or specified business domain, technical skills to create and run analytics model to obtain insight from data, the ability to analyze and interpret data (IIBA, 2019). A. Verma et al. (2019) extend this list by grouping individual competencies of data analysts into the following categories: enterprise systems software, visualization techniques, specialized analytics solutions, programming skills, project management, advanced modeling/analytics techniques, web scraping, hardware, networking, statistical package, data mining techniques, structured data management, big data management, decision making skills, communication skills, organization skills and business domain.

Our research did not focus directly on the competencies of a business analyst, but presented a number of Job Position Responsibility related to the competencies presented in the literature. Examples of keywords of job position responsibility analyzed by us are as follows: process, project, report, data, management, decision making, preparation financial data, internal cooperation, budgeting, controlling, etc. These areas require the competencies described by J. Park and S. Jeong (2016) and A. Verma et al. (2019).

The study also confirms the conclusions of the 2021 Global State of Business Analysis Report. This report presents a list of job titles where employees most often define their role in the organization as a business analyst: business analyst, business analyst/project manager, IT business analyst, consultant, manager/director/VP/C-level, business systems analyst, product owner/product manager, senior business analyst (IIBA, 2021). According to the 2021 Global State of Business Analysis report, the key competences of a business analyst are: communication, problem solving, creative thinking, facilitation, organizational knowledge, leadership & influencing, specialized skills (IIBA, 2021).

5. Summary

The idea of the research was to use ensemble approach of the expert judgment and machine learning methods to identify candidate profiles for analyst job position. The object of the analysis was the sample of job offers published in the internet job offer portal operating in Poland.

As the first, we have analyzed the most demanding skill levels for an analyst job position. As we shown, the most frequently (almost 60%) a mid and regular level of specialists are searched by offerors. The next two positions occupy junior and senior with almost the same (18%, 17% respectively) demand.

In the second part, we used LDA topic modelling from Mellet package to made a job offer's text (job title) summary into 50 topics and by utilizing expert's judgements establish seven categories of analyst jobs based on keywords appearing in job offers. We built the most descriptive vocabulary for every category and analyzed job profiles.

As the next step we analyzed job descriptions (responsibilities). We applied previously established classification, keywords, vocabulary and machine learning to classify jobs according job description. We obtained confusion matrix where we associated job categories for job title and job description. This allows us to deliberate about coherence between job title and job description in context of recruitment marketing and job position understanding in organization.

However our dataset was not representative for any society group or industry the research delivers interesting highlights for academia and practitioners:

a) An application of machine learning text mining techniques for job's offer analysis and categorization, however very rare in literature, is possible and delivers functional results.
b) We combined expert method of judgment with machine learning preprocessing and then utilize computational technique to extend the judgment to wide range sample of data what is essential for supervised machine learning classification.
c) We detected significant inconsistency in job description and job titles advertised in recruitment process. This suggests that managers can mix requirements for different positions with improper use of job title in recruitment process. This means that job title is used as prestige title is some cases where the job description is taken from more common position. We suspect that a number of offers for "financial controller" is functioning as prestige job title for "financial analyst". Similarly "data scientist" is functioning as for other analytical positions (business, financial).

d) Imprecise job description lead also to confusion in relationships on organizational level. If candidate or organization understand job title in some way but recruiter is asking something else or candidate expect different requirements the confusion can occurs in both side.

We believe the presented study is important both from the point of view of science and managerial practice. We are going to continue our work in further research to recognize the profile of candidates and work environment which offerors propose for the new type of workers as hybrid work place, flexible time, freelance employment, gig economy, etc. It will contribute to increasing the state of knowledge about the function of a business analyst in organizations, and at the same time it can be a guide for managers and HR departments looking for competent business analysts.

References

1. Aasheim, C.L., Williams, S., Rutner, P., Gardiner, A. (2015). Data Analytics vs. Data Science: A Study of Similarities and Differences in Undergraduate Programs Based on Course Descriptions. *Journal of Information Systems Education, 26(2),* 103-115.
2. Babar, A., Bunker, D., Gill, A.Q. (2018). Investigating the relationship between business analysts' competency and IS requirements elicitation: a thematic-analysis approach. *Communications of the Association for Information Systems, 42(1),* 12.
3. BAE (2014). *Business Analyst Roles and Responsibilities*, http://business-analysis-excellence.com/business-analyst-roles-and-responsibilities/, 1.03.2023.
4. Barde, B.V., Bainwad, A.M. (2017). *An overview of topic modeling methods and tools*. 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, pp. 745-750.
5. Blei, D. M., Ng, A.Y., Jordan, M.I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3(Jan),* 993-1022.
6. Chakabuda, T.C., Seymour, L.F., Van Der Merwe, F.I. (2014). *Uncovering the competency gap of students employed in business process analyst roles—an employer perspective*. IST-Africa Conference Proceedings. IEEE, pp. 1-9.
7. Chauhan, V.K. (2015). *How to reduce user story reopen count in Scrum development?* 2nd International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, pp. 2204-2208.
8. Chernysheva, Y.G., Shepelenko, G.I. (2018). *The new profession of "business analyst" and the new occupational standards: the case of Russia*.
9. Daneva, M., Herrmann, A., Condori-Fernandez, N., Wang, C. (2019). *Understanding the most in-demand soft skills in requirements engineering practice: Insights from two focus*

groups. Proceedings of the Evaluation and Assessment on Software Engineering, pp. 284-290.
10. Deng, R., Zhong, S. (2022). Evaluation and Dynamics Mechanism of the Smart Specialization Level in China's Regional Manufacturing Industry. *Innovation and Development Policy, 4*, 51-77.
11. Ghosh, A. (2015). Outline of Dynamical Business Analysis Function for Organizational Structure. *International Journal of Scientific & Engineering Research, 6*, 1898-1909.
12. Gobov, D., Maliarcuk, C., Kunanets, N., Oliinyk, Y. (2020). *Approaches for the Concept "Business Analysis" Definition in IT Projects and Frameworks.* ICST, pp. 321-332.
13. Goh, J., Lee, J., Hur, W., Ju, Y. (2019). Do analysts fully reflect information in patents about future earnings? *Sustainability, 11(10),* 2869.
14. Graham, S., Weingart, S., Milligan, I. (2012). *Getting Started with Topic Modeling and MALLET.* UWSpace. http://hdl.handle.net/10012/11751
15. IIBA (2015). *A Guide to the Business Analysis Body of Knowledge (BABOK), International Institute of Business Analysis, Ver3*, https://www.iiba.org/career-resources/a-business-analysis-professionals-foundation-for-success/babok/, 1.03.2023.
16. IIBA (2019). *Introduction to Business Data Analytics: An Organizational View*. Toronto: International Institude of Business Analysis.
17. IIBA (2021). *Global State of Business Analysis in Report*. International Institute of Business Analysis, https://go.iiba.org/global-ba-salary-survey/2021, 1.03.2023.
18. Liu, X., Wang, C. (2021). *An Empirical Study on Hyperparameter Optimization for Fine-Tuning Pre-trained Language Models*. arXiv preprint arXiv:2106.09204.
19. Ndlela, M., Tanner, M. (2022). Business analysts' contributions to the dynamic capabilities of agile software development teams. *Information Technology & People*.
20. Office for National Statistics (UK) (August 1, 2022). *Total numbers of IT business analysts, architects and systems designers in the United Kingdom (UK) from 2011 to 2021 (in 1,000s) [Graph]*. Statista. Retrieved from: https://www-1statista-1com-1j013e2w500da.han.bg.umcs.edu.pl/statistics/318816/numbers-of-it-business-analysts-architects-and-systems-designers-in-the-uk/, 16.04.2023.
21. Park, J., Jeong, S.R. (2016). A study on the relative importance of underlying competencies of business analysts. *KSII Transactions on Internet and Information Systems (TIIS), 10(8),* 3986-4007.
22. Richards, D., Marrone, M. (2014). Identifying the education needs of the business analyst: an Australian study. *Australasian Journal of Information Systems, 18(2).*
23. Shah, M. (2017). Evolving Role of a Business Analyst. *International Journal of Business and Management, 1(2),* 7-12.
24. Verma, A., Yurov, K.M., Lane, P.L., Yurova, Y.V. (2019). An investigation of skill requirements for business and data analytics positions: A content analysis of job advertisements. *Journal of Education for Business, 94(4)*, 243-250.

25. Wang, C., Wu, Q., Weimer, M., Zhu, E. (2021). Flaml: A fast and lightweight automl library. *Proceedings of Machine Learning and Systems, 3,* 434-447.
26. Zych, M. (2020). Education for business analysts in Poland. *Education for Information, 36(4),* 347-370.

Appendix 1

Most frequent key words in selected categories

Category	Most frequent key words	Topic
business analyst	analyst, market, data, power, trading, energy	0
	process, automation, robotics, trainer	1
	analyst, service, business, customer	11
	monitoring, transaction, remediation	14
	wealth, banking, personal, mobile, marketing	22
	analyst, sale, business, support, development, system	24
	analyst, manager, project, operation	32
	analyst, real, estate	33
	analyst, corporate, business, treasury, client, investment	34
	analyst, management, market, risk, quantitative, asset	38
	analyst, business, system, consultant	39
	management, expert, contract, ssc, travel	41
	product, owner, manager	42
	analyst, pricing, supply, transfer, chain, planner	43
	analyst, center, excellence, pay	48
	specialist, controlling, analysis, planning	49
data scientist	data, scientist, engineer, lead	3
	program, graduate	6
	analytics, associate, team, sql	8
	data, innovation, analytical, forecast	15
	data, engineer, developer, cloud	18
	developer, report, sql, etl, power, record, programmer	27
	specialist, analysis, data, reporting	30
	analyst, data	40
financial analyst	analyst, credit, risk, management	2
	excel, resilience, national, hire	5
	research, offshore, black, sea, grain, oilseed	7
	analyst, control, cost, financial, specialist, operation	10
	analyst, tax, accountant, italian	13
	support, financial, valuation	16
	specialist, settlement, billing	17
	account, accountant, receivable, ptp, payable, lease	20
	analyst, reporting, accounting, finance, financial, specialist	23
	analyst, fund, accounting, investment	29
	derivative, insurance, asset, class	47
financial controller	controller, financial, controlling, team, project, officer	9
	financial, analyst, accountant, business, controller	26
	financial, controller, analyst, finance	28
	financial, analyst, accountant, business, controller	36
	accountant, general, ledger, accounting, controller	45
intern analyst	internship, audit, crm, consultant, summer, team	35
	intern, processing, insurance, life	37
security analyst	analyst, security, soc	12
	designer, fraud	25
	client, prevention, anti, know, money, laundering, onboarding	31
	security, revenue	44
	analyst, kyc, aml	46
	sap, center, key, user	4
	performance, analyst, tester	19
	network, administrator, computer	21

Appendix 2

Most frequent key words in job requirements for technology

		in tot %	30,9%	5,6%	9,5%	2,3%	1,4%	2,3%	1,6%	2,5%	1,1%	2,7%	1,9%
	19 094	769	3 049	171	291	71	43	69	50	76	33	82	57
index	**dł**	**count**	**tech_len**	**python**	**sql**	**jira**	**spark**	**uml**	**google**	**azure**	**excel**	**power**	**java**
sql	3	46	1	0	1	0	0	0	0	0	0	0	0
python sql	10	16	2	1	1	0	0	0	0	0	0	0	0
sql python	10	15	2	1	1	0	0	0	0	0	0	0	0
jira confluence	15	12	2	0	0	1	0	0	0	0	0	0	0
sap	3	10	1	0	0	0	0	0	0	0	0	0	0
uml bpmn	8	9	2	0	0	0	0	1	0	0	0	0	0
	0	8	1	0	0	0	0	0	0	0	0	0	0
sql power bi	12	6	3	0	1	0	0	0	0	0	0	1	0
enterprise architect	20	6	2	0	0	0	0	0	0	0	0	0	0
python	6	6	1	1	0	0	0	0	0	0	0	0	0
bpmn confluence jira sql enterprise architect	45	6	6	0	1	1	0	0	0	0	0	0	0
agile scrum	11	5	2	0	0	0	0	0	0	0	0	0	0
m office vba python powerquery power bi	39	5	7	1	0	0	0	0	0	0	0	1	0
active directory	16	4	2	0	0	0	0	0	0	0	0	0	0
uml bpmn sql	12	4	3	0	1	0	0	1	0	0	0	0	0
python panda scikit learn kera pytorch tensorflow aws s	105	4	16	1	1	0	1	0	0	0	0	0	0
sql sa python	13	4	3	1	1	0	0	0	0	0	0	0	0
bpmn uml	8	4	2	0	0	0	0	1	0	0	0	0	0
google analytics tag manager sql python	39	4	6	1	1	0	0	0	1	0	0	0	0
sa	2	3	1	0	0	0	0	0	0	0	0	0	0
agile	5	3	1	0	0	0	0	0	0	0	0	0	0
uml enterprise architect	24	3	3	0	0	0	0	1	0	0	0	0	0
azure	5	3	1	0	0	0	0	0	0	1	0	0	0
google analytics tag manager	28	3	4	0	0	0	0	0	1	0	0	0	0
sparx enterprise architect	26	3	3	0	0	0	0	0	0	0	0	0	0
sql python sa	13	3	3	1	1	0	0	0	0	0	0	0	0
sql python git google cloud	27	3	5	1	1	0	0	0	1	0	0	0	0
python java opencv scikit panda pytorch nlp jenkins dock	79	3	13	1	0	1	0	0	0	0	0	0	1
jira confluence sql	19	3	3	0	1	1	0	0	0	0	0	0	0
microsoft excel	15	3	2	0	0	0	0	0	0	0	1	0	0
sql python google cloud platform	32	2	5	1	1	0	0	0	1	0	0	0	0
uml soapui postman	18	2	3	0	0	0	0	1	0	0	0	0	0
python docker kubernetes	24	2	3	1	0	0	0	0	0	0	0	0	0
sql hive python apache spark	28	2	5	1	1	0	1	0	0	0	0	0	0
confluence jira	15	2	2	0	0	1	0	0	0	0	0	0	0
figma	5	2	1	0	0	0	0	0	0	0	0	0	0
sql bpmn uml microsoft office erp system crm sent emcs	68	2	12	0	1	0	0	1	0	0	0	0	0
agile scrum six sigma	21	2	4	0	0	0	0	0	0	0	0	0	0