Report on Scientometric Analysis of "The role of Big Data in Modern Project Management" [1, 2]

by Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

Problem

Informational technology has been developing very rapidly over the last few decades. Informational technology has become an important part of modern science and industry. The other thing, whose popularity has been growing as an important part of modern industry, is Big Data. Big Data's importance has been proven by its high value in offering analytics and better solutions based on big amounts of quantitative data, which can help to design sophisticated projects with fine decisions.

Research purpose

The goal of this research report is to understand the level of integration of Big Data in the Project Management sector, the trends, correlation and level of linking between different categories in the bonding of Big Data and Project management.

Methodology

To achieve our purpose, the study analysed 773 publications relevant to Big Data, Project Management and its allied categories. A Co-word Analysis Technique (CWAT), a scientometric method for quantitative analysis of the content of academic publications, was applied for this study. The study utilised the VOSViewer® software tool specifically designed for constructing and visualising bibliometric maps. Our keywords were: Big Data, Project, Management, Role.

Scopus search settings

The search data on the Scopus portal was with the limitations to be relevant to the Master's Program research. The search keywords look in a such way:

T*ITLE-ABS-KEY ("big data" AND "project" AND "management" OR "role") AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (DOCTYPE , "ar") OR (LIMIT-TO (DOCTYPE , "re")) AND (LIMIT-TO (LANGUAGE , "English"))*

[1] How to cite this article: Zhambyl, A., Abdeshova, A., Kenzhibayeva, Z. (2022). Report on Scientometric Analysis of "The role of Big Data in Modern Project Management", *PM World Journal*, Vol. XI, Issue VIII, August.

[2] Editor's note: This paper was produced for the course Business Research Methods in the Project Management graduate program at the Kazak British Technical University in Kazakhstan. Supervising professor was Dr. Timur Narbaev.

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

Search link: https://www.scopus.com/results/results.uri?sort=plf-f&src=s&sid=1869cedb47e57ca78f651f53612590a3&sot=a&sdt=a&cluster=scosrctype%2c%22j%22%2ct%2bscosubtype%2c%22ar%22%2ct%2c%22re%22%2ct%2bscolang%2c%22English%22%2ct&sl=69&s=TITLE-ABS-KEY+%28+%22big+data%22+AND+%22project%22+AND+%22management%22+OR+%22role%22+%29&origin=searchadvanced&editSaveSearch=&txGid=32420709e55b7be1691303f3c2c4f42c

The main keywords are Big Data, Project, Management and Role. The source type was chosen to be limited to journals and the document type is article and review. Only the sources in the English language were considered for this research.

Export data

For the export of the Scopus database results it was chosen to export using the CSV file format, which can be opened using the Microsoft Excel program and imported by VOSViewer®.

The settings for the export data are such (see Table 1. CSV export):

Export type	**CSV file**
Citation information	Document title
	Year
	Source title
	Citation count
	Source & document type
Bibliographical information	(no point was chosen)
Abstract & keywords	Abstract
	Author keywords
	Index keywords
Funding details	(no point was chosen)
Other information	(no point was chosen)

Table 1. CSV export

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

VOSViewer® settings

VOSViewer® is a powerful tool, which can fetch the data from the Scopus database, where the articles were monitored. The software makes a quantitative analysis of the data based on the number of occurrences, relevance, co-appearance in the similar articles and links between different groups.

The software only makes quantitative analysis and does not provide detailed information on how to understand the outcomes. The research team makes the interpretation of the outcome and gives some insights from the visualised result in a form of trends, comparison and pattern using the tables.

The settings for the bibliometric analysis are such (see Table 2. VOSViewer® initial settings):

Number of occurrences	10
Type of analysis	Co-occurrence
Counting method	Full counting
Number of keywords to be selected	113

Table 2. VOSViewer® initial settings

After the simulation run, the research team made a few adjustments for the data visualisation. The settings for the visualisation are such (see Table 3. VOSViewer® settings adjustments):

Minimum cluster size	Changed from 1 to 22 (resulted in change of cluster down to 3 clusters)

Table 3. VOSViewer® settings adjustments

Words exclusion

The research team used the option to exclude the common and inappropriate words from the visualised data in order to keep the relevance of the research. The words from the file thesaurus.txt were excluded for the visualisation.

Results and findings

Three clusters. There are 3 clusters, which give us visualisation of the data.for the 164 most occurring terms (see Figure 1). The distribution of the items can be distinguished

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

by the relativity to its cluster, which is identified by its colour. On the visualised data, there are three colours, which represent separate clusters (see Table 4. Main 3 clusters).

Cluster number	Cluster name	Cluster colour
1	Big Data in Project Management	Red
2	People and Health related data research	Green
3	Data analysis and Information processing	Blue

Table 4. Main 3 clusters

Cluster 1 - Big Data in Project Management (in red colour). This cluster focuses on areas affected by Big Data and data and methods are used to process the information. In the Cluster 1, the areas where Big Data is used are applied, such as Machine learning, information management, project management, decision support system, construction, forecasting, risk management, urban planning, artificial intelligence, quality control. This cluster includes the terms (see Table 5. Cluster 1):

Item name	Occurence times	Occurence total link strength
Information management	97	381
Project management	70	279
Data handling	42	208
Artificial intelligence	45	218
Decision making	56	238

Table 5. Cluster 1

Interpretation of the Cluster 1. The "Information management" is the item which appeared most and has the second strongest link. The "Project management" item appears the second most appeared item in the "Big Data in Project Management" cluster. There can be a trend that the stronger the occurrence, the stronger the link, however there are some exceptions that items with less occurrence have stronger link like "Data mining" items.

Cluster 2 - People and Health related data research (in green colour). This cluster focuses on human interaction-based areas. The second cluster implies personalised data. This cluster mostly focuses on biomedical research, medical research, innovation, algorithms, standards and procedures, semantics, bioinformatics, software. Also, many groups refer to medical and public health related areas. This cluster includes the terms (see Table 6. Cluster 2):

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

Item name	Occurence times	Occurence total link strength
Human	155	975
Procedures	40	381
Priority journal	35	257
Software	26	203
Biomedical research	22	185
Medical research	25	212

Table 6. Cluster 2

Interpretation of the Cluster 2. The item "Human" has appeared the most and visually it can be seen that most of the bigger items in the Cluster 2 are related to health. Here, it can be understood that the occurrence times correlate with the occurrence link strength.

Cluster 3 - Data analysis and Information processing (in blue colour). This cluster focuses on human related medical records. This cluster is more general compared to cluster 2, which has more personalised groups and intersects with the categories such as government, computer security and privacy. Overall, clusters 2 and 3 are similar. This cluster includes the terms (see Table 7. Cluster 3):

Item name	Occurence times	Occurence total link strength
Information processing	37	326
Data analysis	41	320
Healthcare	43	331
Data mining	65	402
Organisation and management	28	237

Table 7. Cluster 3

Interpretation of the Cluster 3. It can be seen from the Cluster 3 that it has similar elements as in Cluster 2 and it is related to Health. However, the Cluster 3 mostly has items related to the processing and analysis of the information. Compared to cluster 2, in cluster 3 we can see the various connections innovations in data processing on a large scale have with the medical field. The processing of big data in order to obtain analytical information can be the next step in providing a single digital contour in the healthcare sector. Now work is underway to provide infrastructure and a basic array of information that can be analysed in the future.

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

Conclusion. Thus, based on the above, since big data directly works with individuals' personalised information, approaches to solving cybersecurity problems to protect enterprises and organisations from threats and attacks also need to be organised for big data, by intelligently managing network traffic and the behaviour of packets transmitted over networks. Usually, in business intelligence solutions, the emphasis is on the use of internal structured data and analytical information based on certain rules, and only then this information is processed to make a particular decision. As follows from the definition itself, the transition to big data leads to an increase in the range and diversity of information entering the management system. In modern systems, this data can be analysed and, therefore, additional information about contributory factors will be obtained.

Network visualisation. It can be seen from the Figure, that there are 3 main clusters, where second and third clusters have closely related items (see Figure 1. Network visualisation).

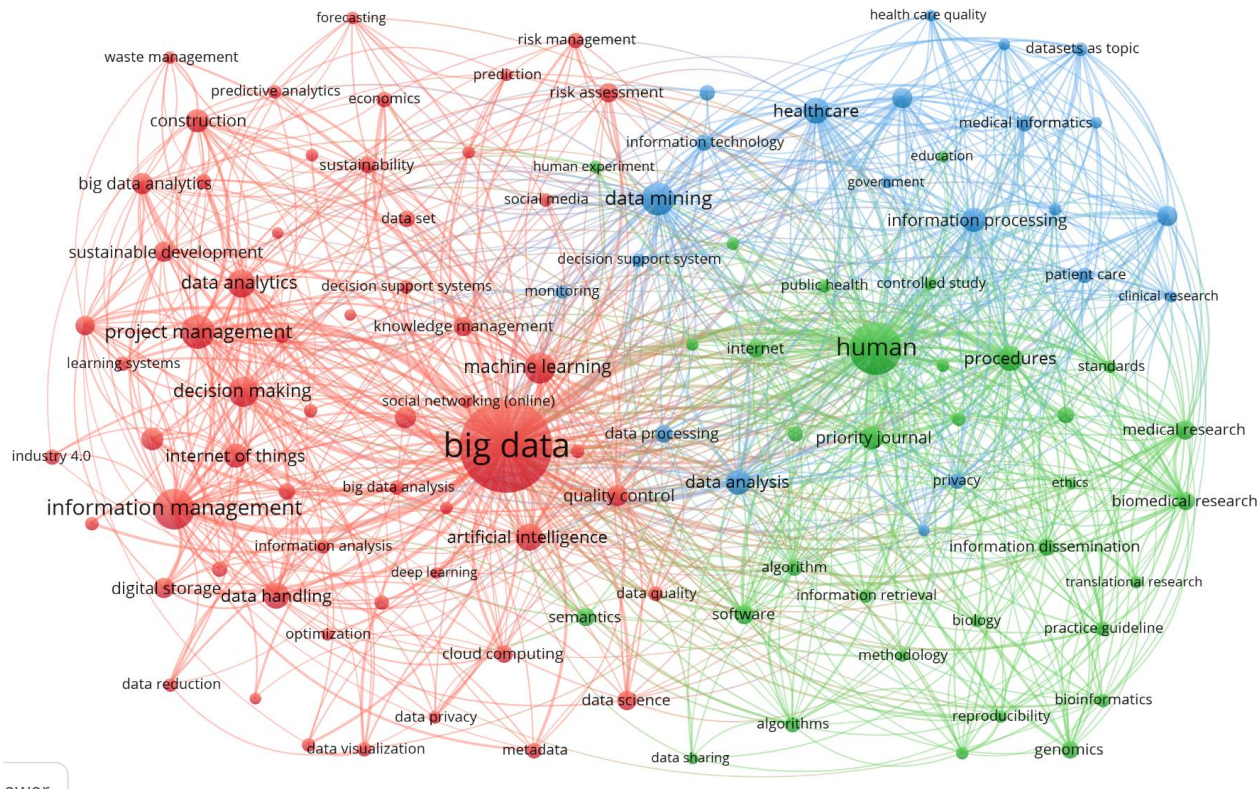

Figure 1. Network visualisation

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

Overlay visualisation. This visualisation demonstrates the heat map of the items' occurrence depending on the articles' publishing date. It can be seen that most of the articles are published between the years 2017 and 2019 (see Figure 2. Overlay visualisation).

For year 2017, the following topics were mentioned more:
- Information processing
- Healthcare
- Clinical research
- Data analysis

For 2018, the following topics were mentioned more:
- Big Data
- Human
- Data mining
- Information management

For 2019, the following topics were mentioned more:
- Decision making
- Machine learning
- Data analytics

For 2020, the following topics were mentioned more:
- Data science
- Predictive analytics
- Artificial intelligence

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

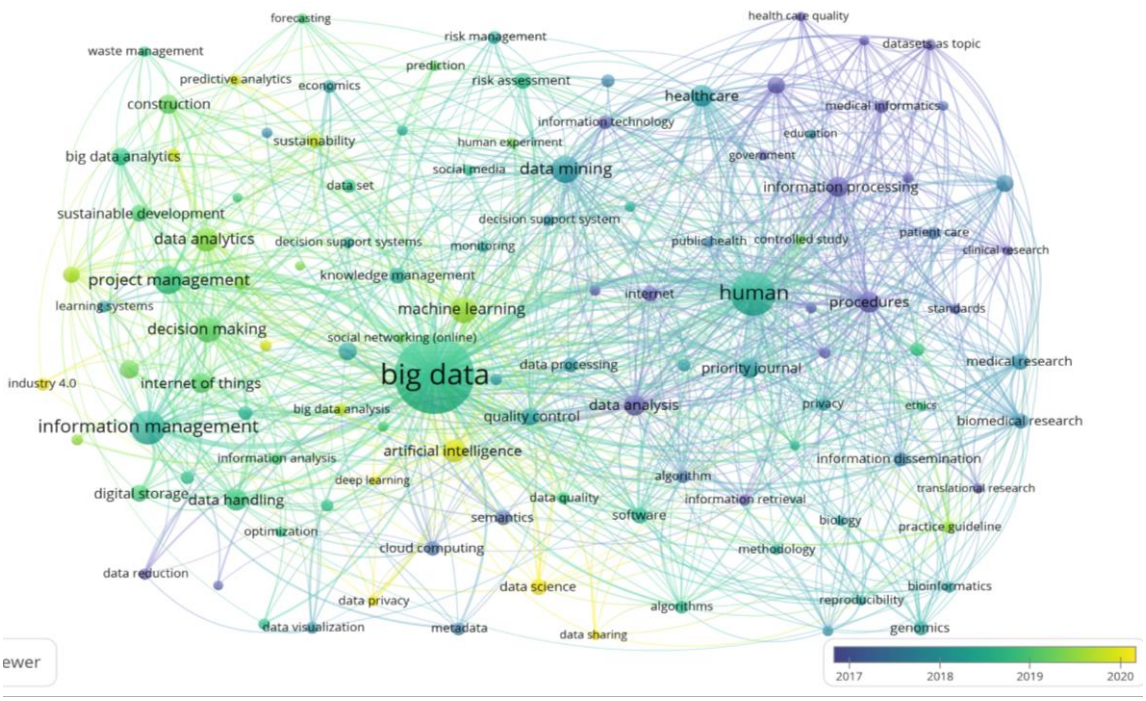

Figure 2. Overlay visualisation

Density visualisation. The visualisation image shows the most densely mentioned items. The most noticeable items are such (see Table 8. Dense items and Figure 3. Density visualisation):

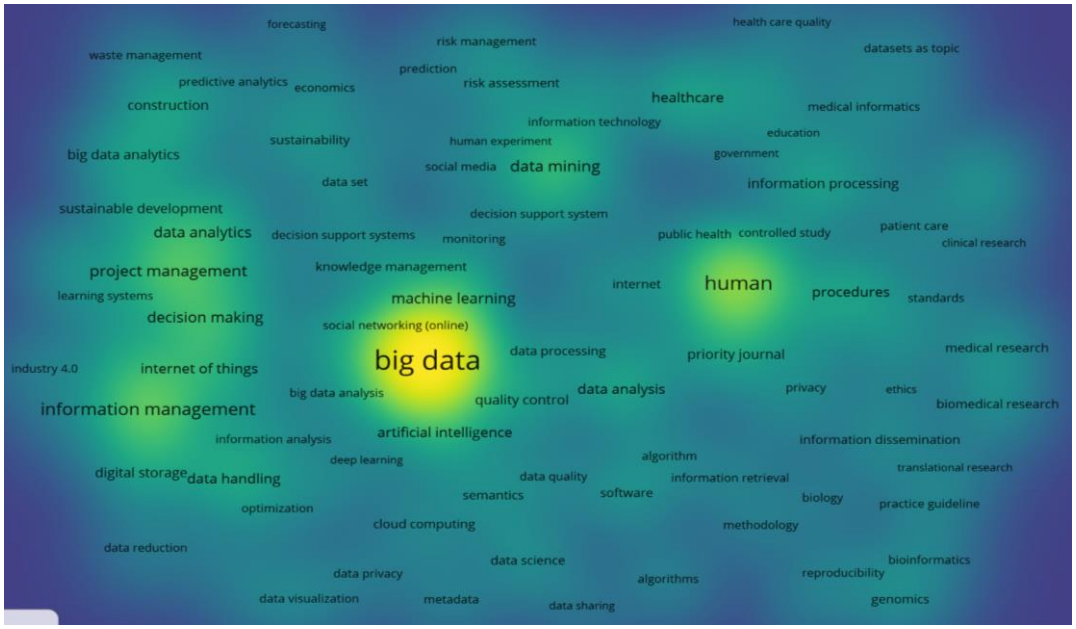

Figure 3. Density visualisation

Item number	Item name
1	Big Data
2	Human
3	Information management
4	Project management
5	Data mining

Table 8. Dense items

Table of most popular items. The software gives an opportunity to view the most popular items based on the total link strength. When considering the clusters separately, there is a logical increase in the number of occurrences as the total link strength grows. However, the table below shows that there is no correlation between the number of occurrences and the total link strength when considering three cluster elements altogether (see Table 9. Most popular items).

Item name	Occurrences	Total link strength
big data	444	1634
human	155	975
procedures	40	381
data mining	65	402
information management	97	381
information processing	37	363
data analysis	41	320
machine learning	58	305
healthcare	43	331
priority journal	35	287
data analytics	50	325
electronic health record	26	281
project management	70	300
organisation and management	28	257
decision making	56	265

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

artificial intelligence	45	236
medical research	25	231
electronic health records	28	281
software	26	225

Table 9. Most popular items

Project Management connections. As this study was done by Project Management students, it is important to analyse the connections of "Project Management" items with the other items. It was decided to choose the relevance of the other elements depending on the distance of different items from the "Project Management" item (see Figure). It can be seen that there are many items from Cluster 1 (red), the few items from other clusters are such (see Table 10. Project Management connections and Figure 4. Project Management connections):

Item	Cluster belonging	Link strength
software	Cluster 2 - People and Health related data (green)	4
semantics	Cluster 2 - People and Health related data (green)	4
data analysis	Cluster 3 - Data analysis and Information processing (blue)	7
information technology	Cluster 3 - Data analysis and Information processing (blue)	4

Table 10. Project Management connections

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

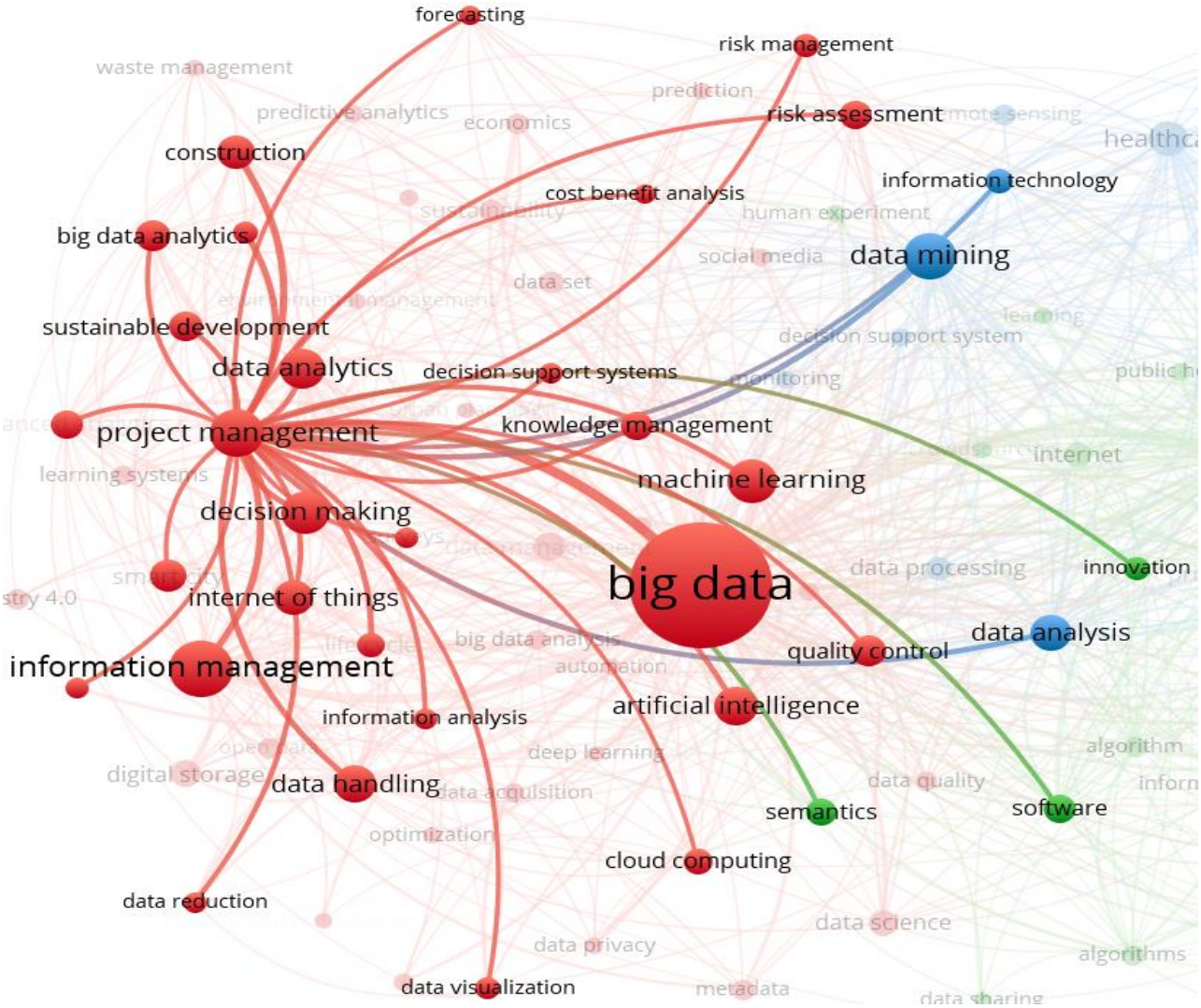

Figure 4. Project Management connections

Discussion

How is Big Data valuable for Project Management? The analysis done before, identifies that the Project Management intersects with other clusters, where the health-related categories are mentioned. The 2 important closer links to Project Management are related to **Construction** and **Decision making**.

Suggestion for the current time. It can be concluded that currently, project managers should focus on benefitting from usage of Big Data while working with Construction to manage the flow of information. The second area is Decision making, where the Project Managers must take the best from the information and data flow by analysing the data and extracting the most important key points, then using that extraction in further plan adjustment.

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

Suggestion for the future for the Project Managers. It is suggested to involve the Project Managers in data analytics and learning the programming languages and tools, which might come handy for more refined decisions and project plan adjustment. The clusters number 2 and 3 show lots of interactions with health-related categories. Thus, it can be suggested that in the future we can witness the growing trend in the usage of Big Data and Data processing in healthcare projects.

About the Authors

Aibek Zhambyl

Kazakhstan

Aibek Zhambyl graduated Kazakhstan Maritime Academy with a bachelor's degree in Marine Engineering and Technologies. Currently working at the Maritime Academy as an assistant lecturer and pursuing a master's degree in Project Management at KBTU to apply the knowledge on big projects in the maritime industry of Kazakhstan. Aibek can be contacted at aibekzhambylmail@gmail.com

Aidana Abdeshova

Kazakhstan

Aidana Abdeshova graduated Satbayev Kazakh National Technical University with a bachelor of science in Mathematics and Economics. Presently enrolled in a master's program in Project Management at KBTU in hopes of gaining knowledge to embark on a fulfilling career in business. Aidana can be contacted at abdeshovaaidana1@gmail.com

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva

Zhanel Kenzhibayeva

Kazakhstan

Zhanel Kenzhibayeva has a bachelor's degree in Radio engineering and telecommunications. This led to her working in the telco industry for 5 years. During two years she worked as a product manager in a mobile operator company. She decided to continue her education and pursue a Master of Science in Project Management from Kazakh-British technical University (KBTU) to learn more about project management. Zhanel can be contacted at kenzhibaeva1997@gmail.com

© 2022 Aibek Zhambyl, Aidana Abdeshova, Zhanel Kenzhibayeva