International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 59-63
Published Online October 2020 in IJEAST (http://www.ijeast.com)

USING DATA SCIENCE IN PUBLIC SECTOR FOR IDENTIFYING SUCCESSOR AND PREDICTING ATTRITION

Nikmohd Abdrashid
IT Appl/Sys Consultant,
Corporate Applications Department
Saudi Aramco, Dhahran,
Eastern Province, Saudi Arabia

Shawqi Bazbooz
IT System Analyst II,
Corporate Applications Department
Saudi Aramco, Dhahran,
Eastern Province, Saudi Arabia

Mohamed Sowaij
IT System Analyst I, Corporate Applications Department
Saudi Aramco, Dhahran, Eastern Province, Saudi Arabia

***Abstract*—Analytics delivers critical insights about employees, their preferences, what makes them most effective, their contribution to business success; and can answer some critical questions across talent lifecycle, including attrition and succession. There is a business need to build predictive models to be able to predict attrition, and identify the best successors to maintain a healthy internal talent pipeline, and retain the best talent to succeed as a high performing organization.**

Keywords— **Data Science, Succession, Attrition, Talent, Prediction**

I. INTRODUCTION

Organizations across the globe are faced with an alarming reality. In the very near future, a significant number of senior leaders will hit average retirement age. How can the organization ensure that this new generation of leaders will be sensitive to its corporate culture and mission, possess the necessary subject matter expertise, and foster appropriate contacts and networks [3]? The absence of an effective succession plan and predicting attrition in an organization can have unforeseen consequences.

Succession planning and predicting attrition are critical factors in the long-term success of the organizations. Whether the changes are planned or not, if they do not have an effective succession strategy in place, the business continuity could be at risk [1]. While most organizations execute the traditional processes for Succession Planning, these often fail to understand the real way in which organizations function. Traditional Succession Planning relies heavily on a manager's subjective evaluation and knowledge of his or her direct reports. They often identify key talent based on the formal organizational hierarchy, contextual HR information, or simply on superficial observations. These approaches ignore the informal organization, the networks that employees form across functions and divisions to get things done [7]. When employees leave, they depart with more than what they know — they also leave with critical knowledge about who they know. Although these networks and relationships are seemingly invisible, losing them poses a substantial threat to an organization. Therefore, an additional network perspective can help managers gain a 360-degree view of their Succession Planning processes [7].

Information, relationships and networks created and maintained by employees, are a rich source of data, which is very important in the field of data science to support better decision making towards selecting the best successor and predicting attrition. The present article will investigate the application of data science to the analytics solution, to identify best successor and predict attrition, how data is gathered, cleaned and explored, how machine learning is used to create predictive modeling, and how the results are communicated.

II. BUSINESS CASE JUSTIFICATION

Organizations are now moving from descriptive analytics that use a historical representation of data to predictive and prescriptive analytics that leverage real-time data and provide futuristic insights through cognitive computing and machine learning. Two important analytics use cases for organizations are succession planning and attrition, as they are essential for a healthy organization to ensure that its new generation of leaders will be sensitive to its corporate culture and mission, possess the necessary subject matter expertise, and foster

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 59-63
Published Online October 2020 in IJEAST (http://www.ijeast.com)

appropriate contacts and networks [3]. The purpose is to use data science to develop analytics models to identify best successor and predict attrition, along with an analytics framework that enables custom modeling for many HR analytics use cases. Data analytics techniques also will help organizations to identify bad trends in managing employees' performance which will hide good talents from being visible[10]. The analytics framework would provide the following:

- Availability of real time analytics that provides a forward looking view of the organization.
- A single source of truth for enterprise wide data, including internal and external, to ensure data accuracy and consistency. The data needs to be stored and processed in a consolidated place where it can be accessed for performing analytics.
- Machine learning capabilities with the use of cognitive computing to provide the most relevant insights.
- Ability to permit scenarios and predictions with flexibility in data modeling, based on the user's requirements.

The goal is to have a strong solution with deep analytic capabilities that can provide predictive and prescriptive insights, putting data at the center of decision making, including identifying successor and predicting attrition.

III. DATA COLLECTION

The technologies and networks used by the organizations are expanding, so does the amount of data available in different fields [2]. Data has been increasing in 2020, and will continue increasing in the coming years. According to the IDC report, the amount of data worldwide will increase from 33 Zettabytes in 2018 to 175 Zettabytes in 2025 [8].

Big companies with large populations of employees are producing huge amount of data and increasing each year in various fields. The necessity to understand and explore this amount of data gave importance to the field of Data science, which can be defined as the scientific approach used to derive knowledge and insight from structured and unstructured data sets, to provide support for decision making [2].

The foundation of this approach is to create a data lake that will be used to create powerful models, capable of providing predictive and prescriptive analytical solutions for all analytics use cases, including identifying best successors and predicting attrition. The data lake should acquire, blend, integrate, and converge all types of data, regardless of sources and format, serve any volume and type. It should include a variety of data from internal and external sources as following:

- Internal data from all HR systems, including: general employee information, organization data, recruitment, training and performance, talent assessment, security system photographs, etc.
- External data like benchmarking data, and social media data.
- Data like GPA, school, patents, publications, coaching history, experience history, full time and temporarily assignment, historical mobility and development opportunities.
- Loan data, medical record, and grievance data.
- Data for position profiles.
- Unstructured Data: data from emails, chats, engagement surveys, uncategorized data in the form of spreadsheets, presentations and text that may be available in the organization.
- Data from external sources like market trends, current and forecasted skill requirements, industry benchmarks like salary, etc.
- Data from point solutions like collaboration platform, chatbot, etc.

There are several compliance laws to consider when collecting data. These are heavily governed and should be kept in mind when implementing any analytics solution. Some of these regulations will concern issues like:

- Employee privacy and anonymity.
- Employee consent for the amount and type of data collected.
- IT security when using third-party software for HR analytics.
- Where the HR data is stored.
- Compliance with local laws.

So it's always wise to collaborate with the company's legal team to ensure compliance with any ethics and statutory regulations [9].

Finally, the collected data needs to be based on SQL Server database and of a single source of truth, to avoid data mismatch across different dashboards

IV. DATA CLEANING AND EXPLORATORY

Cleaning and preparing the data is the most time consuming step of all, and this is especially true in big data projects, like identifying successor and predicting attrition, which involve terabytes of data. According to interviews with data scientists, this process can often take 50 to 80 percent of their time [4]. The reason why this is such a time consuming process is simply that there are so many possible scenarios and inconsistencies that could necessitate cleaning, and it's important to catch and fix them in this stage [4]. One of the

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 59-63
Published Online October 2020 in IJEAST (http://www.ijeast.com)

steps that are often forgotten in this stage, causing a lot of problems later on, is the presence of missing data. Missing data can throw a lot of errors in machine learning model creation [4].

Once the data is available, cleaned and prepared, analysis gets started. The data exploration is to generate insights at every level (organization, department and individual) and run comparative analysis to make data-driven decisions toward the analytics solution, in this case identifying the best successor and predicting attrition. For examples, analysis of the relationship between high potential employee and the position profile to identify a possible successor, and between employee medical records and his/her type of work to predict possible attrition.

Data scientists and different stakeholders, such as management and professional development department and personnel department, need to meet regularly and discuss, analyze and explore all related data, to come up with possible informative features, data patterns, and contributing factors. They also discuss various types of metrics that can be derived from the collected data [11]. The output of these discussions will help in the design of the analytics solution model, to identify best successor and predict attrition.

Below are some points of discussions for identifying the potential successor:

- Who are the high performers?
- What are the common characteristics of high performers in a role?
- What are the critical roles?
- What skills would be required in the future?
- What is the career path of a high performer?
- What are the development needs for potential successors?
- What are the bench strengths for a successor?

Below are some points of discussions for predicting attrition:

- What percent of the workforce will be reduced?
- Who will leave in the future?
- When will employees leave?
- Why did the ex-employees leave?
- What is the cost of losing high performers?

Numerous research have concluded that the best model depends on the richness of the collected data to accurately identify the best successor and accurately predict attrition [12].

V. MODELING

Machine learning is the process of using domain knowledge to transform the raw data into informative features that represent the business problem to be solved. These features directly influence the accuracy of the predictive model that will be constructed at the end. New features can also be constructed from merging multiple ones to make them more informative by taking their sum, difference or product [4]. For example to identify successor for a position, possible features to consider are readiness and ranking. Other example to predict attrition is to consider features like age and medical record.

Now, it is the time to work for the model based on machine learning that can access and observe data such as examples, direct experience, or instruction, to look for patterns in data and make better decisions in the future, based on the examples that are provided. Note that the task is not just to train the model over accuracy, but also to use comprehensive statistical methods and tests, to ensure that the outcomes from the model actually make sense and are significant. The model should have a solution with analytic capabilities that can provide predictive and prescriptive insights, putting the informative features at the center of decision making.

Below are some analytics to be considered when working for the model to identifying successors:

- Predict and assess available talent pipeline for critical roles, and readiness potential vs. vacancy timeframes.
- Run profile matches of employees against role requirement data to identify potential successors.
- Identify and bridge any skill and competency gaps.
- Identify overall talent needs for development and hiring.
- Identify and predict key competencies for critical roles.

Below are some analytics to be considered when working for the model to predict attrition:

- Forecast attrition of high performers and identify upcoming vacancies due to potential voluntary retirements.
- Forecast percentage of employees leaving in future.
- Predict the most probable leavers and generate risk alerts for preventive actions at the individual level.
- Predict the retention factors for the new generation of employees. [13]
- Estimate the time to attrition for employees for proactive action planning.
- Utilize the exit surveys/interviews, along with other data sources to identify key drivers of attrition.

- Use collaboration platform and chatbot data to better understand employee sentiments and it's impact on attrition.
- Perform a cost benefit analysis for high performer retention.

Undoubtedly, the most critical part of the data science process is to have strong, flexible and robust model that can process all the available data and produce the near accurate insights that is required by top administrators in the public sector.

VI. COMMUNICATE THE RESULTS

After finalizing the models of identifying successor and predicting attrition in real time, they need to be exposed with an open API interface. The interface enables the models to be easily consumed from various applications. The next step is to represent the results in a way that the different key stakeholders can understand. This is called data visualization, which combines the fields of communication, psychology, statistics, and art, with an ultimate goal of communicating the data in a simple yet effective and visually pleasing way [4]. It is also important to ensure management trust the data quality of the models and the actionable insights they can generate.
Below are some analytics to identify successors that might be visualized by the system:

- The key successor pool for each critical role.
- Current and future critical role requirements.
- Planning and hiring decisions based on identified role specific requirements to bridge competency gaps.
- Critically important employees to be highlighted as the successor.
- Employee development programs to align with predicted skill requirements.
- Risk assessments of vacancy for critical roles.

Below are some analytics for predicting attrition that might be visualized by the system:

- Flight risk at individual level of critical roles.
- Proactive inputs for talent acquisition for backfilling of hard to hire skills.
- Talent pipeline to ensure business continuity.
- Comparative attrition analysis by location, demographics, job levels, performance, etc.

Once the system is ready, set of deliveries need to be provided, such as a status dashboard that displays the system health and key metrics, a final modeling report with deployment details and final solution architecture document. Finalized project deliverables confirm that the pipeline, the models, and their deployment in a production environment satisfy the customer's objectives. In addition the business continuity plan has to be defined to include a procedure that will be followed if the system reached a critical situation. Also relevant information regarding information security has to be provided.

The system then will be piloted to provide a platform for the organization to test logistics, prove value and reveal deficiencies before spending a significant amount of time and energy on a large-scale population. After the completion of the pilot phase, the model will be adjusted and enhanced, and the overall solution will be scaled up at the enterprise level.

VII. CONCLUSION

Companies that adopt contemporary data science analysis can expect business expansion and higher revenues. For instance, data scientists calculated that the U.S. healthcare system could save $300 billion annually through an adequate implementation of data science principles [6]. The implementations of data science use cases require very skillful data scientists who possess mastery of machine learning, analytics, statistics and data visualization.

This article discussed two important data science use cases for company's growth; Identifying best successor and predicting attrition as they execute smooth transition of critical roles and leadership positions. It highlighted the challenges in each stage of implementing the analytics solution to identify best successor and predict attrition, and gave examples and recommendations to overcome these challenges. These recommendations are the results of our experience at Saudi Aramco in succession planning and analytics solutions. We also learned from the experience of other leading companies and the best practices followed worldwide.

The article also mentioned and explained the need of an analytics framework that would enable custom modeling for many HR analytics use cases, as data science is a must in the turbulent and chaotic digital world of data. The overall HR functions would be based on scientific and real-time evidence to support the quest to thrive and succeed [2].

VIII. ACKNOWLEDGEMENT

The authors would like to thank the management of Saudi Aramco for their support and Permission to publish this article.

IX. REFERENCE

[1] Digital HR (2020), "Top 10 Succession Planning Tools and Software", [Online].Available, (pp 1-9). https://www.humanresourcestoday.com/data/succession-planning/

[2] Zerktouni, Jabrane (2018), "Data Science, A Necessity For Hr In The Competitive Business World", [Online].Available. https://www.analyticsinsight.net/data-science-a-necessity-for-hr-in-the-competitive-business-world/

[3] Brockbank, Thom and Turi, Ed (2018), "Four Steps for Using AI and Machine Learning for Succession Planning". [Online].Available, (pp 1-2) https://blogs.oracle.com/profit/four-steps-for-using-ai-and-machine-learning-for-succession-planning

[4] Sanat, (2019) "Data Science Life Cycle 101 for Dummies like Me", [Online].Available, (pp5-7) https://towardsdatascience.com/data-science-life-cycle-101-for-dummies-like-me-e66b47ad8d8f

[5] Shao,Cecelia (2018), "A Data Scientist's Guide to Communicating Results". [Online].Available, (pp 2-3). https://medium.com/comet-ml/a-data-scientists-guide-to-communicating-results-c79a5ef3e9f1

[6] Richadson, Chris (2017), "The Most Important Lessons learned from Data Science Projects" [Online].Available, (pp1-4) https://dzone.com/articles/the-most-important-lessons-learned-from-data-scien

[7] People Analytics (2016), "Succession Planning: Why Companies Need More Data To Get It Right." [Online].Available, (pp 2-4). https://www.trustsphere.com/wp-content/uploads/2017/06/Succession-Planning-201610-1-A4.pdf

[8] IDC (201), "The Digitization for the world From Edge to Core." [Online].Available, (pp 3-5). https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

[9] Zander, Olle (2020) ,"Why You Need to Start Working With HR Analytics" [Online].Available, (pp16-18). https://www.trustcruit.com/blog/why-you-need-hr-analytics/

[10] Data Clear. (2013), " How Data Analytics Can Support Succession Planning". [Online].Available, (pp 1-2). http://www.datacleargroup.com/how-data-analytics-can-support-succession-planning/

[11] Mainkar, Gautam. (2017). "Succession planning analytics: Leveraging HR data to ensure stability in the organization" [Online].Available, (pp 1-2) https://blog.harbinger-systems.com/2017/09/succession-planning-analytics-leveraging-hr-data-to-ensure-stability-in-the-organization/

[12] Khot, Veer. (2019) "Using Data Science to Predict Attrition: Retaining people using Artificial Intelligence", [Online].Available, (pp 9-10). https://medium.com/up-engineering/attrition-8a716982a7a8

[13] Rise (2019) "8 Essential Employee Retention Factors Modern Employers Ignore", [Online].Available, (pp 3-4). https://risepeople.com/blog/employee-retention-factors/