

Who are the Leaders in the Data Mining Industry?

PETR BERKA

University of Finance and Administration, Prague, Czech Republic

Abstract: Gartner Inc, a world's leading research and advisory company has developed a methodology how to identify leading companies in a particular market segment. The so called Magic Quadrant uses two complex criteria, "completeness of vision" and "ability to execute" to group possible vendors into four groups: leaders, visionaries, challengers and niche players. Based on the Gartner reports, the paper shows the leading companies in the data mining industry in the last five years. The corresponding data mining systems are briefly introduced as well.

Keywords: magic quadrant; data mining systems.

1 Introduction

The markets for various products and services develop in similar ways. At the beginning, when a market emerges, e.g. due to a new technology, a number of companies appear offering new partial products or solutions that still have a limited impact. In the next phase, high growth can be observed when new vendors enter the market as they are attracted by its potential. Then, a consolidation phase follows, when the growth of the market is not as high but still there is space for some new vendors entering it. When the market becomes mature, only a few key players remain in established positions.

To assess the position of vendors on the market during the high growth and consolidating phases of the market development, Gartner Inc., a leading consulting company, designed a method called Magic Quadrant. Based on two compound criteria, Magic Quadrant visualizes the vendors as divided into four groups: the Leaders, the Challengers, the Visionaries and the Niche Players.

The rest of the paper is organized as follows: Section 2 describes the Magic Quadrant in general, Section 3 discusses the Magic Quadrant for data science for the year 2018, Section 4 focuses on vendors identified as the leaders in 2018 and shows how their position has changed in the last five years. This section also presents data mining tools of the leaders on the market. Section 5 concludes the paper.

2 Gartner Magic Quadrant

Gartner Inc. is a world's leading research and advisory company in the areas of IT, Finance, HR, Customer Service and Support, Legal and Compliance, Marketing, Sales, and Supply Chain. Gartner provides their clients with insights and analyses of various aspects of business and industry. Among them, the Magic Quadrant is a popular and widely used method to compare vendors of some products or services.

Following the idiom "One image counts for thousand words", Magic Quadrant is a graphical representation of the position of the selected vendors on the market. The basics of the Magic Quadrant are two compound criteria: completeness of vision and ability to execute.

Completeness of vision is composed of market understanding, marketing strategy, sales strategy, offering product strategy, business model, industry strategy, innovation and geographic strategy components. Completeness of vision thus refers to the strategies of developing innovative products or services. Ability to execute is composed of products/services, overall viability, sales execution/pricing, market responsiveness, marketing execution, customer experience and operations components. Ability to execute thus refers to the practical issues related to sales and marketing [5]. So, the criteria summarize the potential of a vendor to develop and successfully deliver innovative products. Based on these criteria, the vendors are divided into four groups. Leaders are vendors that have high rating in both the completeness of vision and the ability of execute; they provide mature products and services that meet market demand, and they can affect its overall direction. Challengers are vendors that have high rating in the ability to execute but low rating in the completeness of vision; they typically have a significant size and resources but may lack a strong vision or innovation. Visionaries are vendors that have high rating in the completeness of vision but low rating in the ability to execute; they have a strong view of how the market will evolve but a lower capability to satisfy all possible customers. Niche players have low rating in both the completeness of vision and the ability to execute; they perform well only in a segment of the market because they typically focus on a particular functionality or geographic region. The categorization of the vendors may vary over time; vendors may move between leaders and challengers as the market may shift; visionaries may become leaders or challengers if they strengthen their ability to attract buyers and keep or lose their innovation potential.

Both criteria are numeric, so a vendor can be displayed as a point in a two-dimensional graph, where the x-axis corresponds to the completeness of vision criterion and the y-axis corresponds to the ability to execute criterion. Magic Quadrant visualizes the vendors divided into four groups (quadrants): the Leaders are positioned in the upper right quadrant, the Challengers are positioned in the upper left quadrant, the Visionaries are positioned in the lower right quadrant and the Niche Players are positioned in the lower left quadrant.

It is necessary to note that not every vendor on the market is included in the Magic Quadrant. Gartner makes some selection focusing only on vendors that are the most important or best suited for the buyers in the market. So, just “being included in the Magic Quadrant” is like “being nominated” for an award.

3 Magic Quadrant for Data Mining and Data Science

The market with tools and systems for data mining seems to fulfil the requirements to be analyzed by the Magic Quadrant technique. Although the area has been known for decades and the first data mining suites were made available in the 1990s, which can mean that the market is already mature (consider the pioneering system Clementine by ISL, Inc. that introduced as early as in 1994 a visual interface allowing to use statistical and DM algorithms in an intuitive way without programming), recently we could observe a new breakthrough in this area that has its expression in new terms like deep learning, big data or data science. And indeed, Gartner Inc. issues Magic Quadrants also for this area and reflects the new trends by changing the names for the reports. So, in 2014-2016, the Magic Quadrant was used for advanced analytics platforms [7], [2], [3], in 2017 for data science platforms [4] and in 2018 for data science and machine learning platforms [5].

Fig. 1 shows the Magic Quadrant for the year 2018, that is the Magic Quadrant for data science and machine leaning platforms [5]. While we can find about 90 commercial and about

25 open source data mining suites on the market (see e.g. <http://www.kdnuggets.com>), the Magic Quadrant contains only 16 vendors. Alteryx, SAS, RapidMiner, KNIME and H2o.ai are the leaders; MathWorks and TIBCO are the challengers; Dataiku, Domino, Microsoft, IBM and Databricks are the visionaries, and Teradata, Anaconda, SAP and Angoss are the niche players. We will focus only on the leader's quadrant in the next section.



Fig. 1 Magic Quadrant for data science for 2018 [5]

4 Leaders in the Data Mining Industry

Gartner Inc. identified five vendors to be the leaders in data mining and data science for the year 2018: Alteryx, SAS, Rapid Miner, KNIME and H2O.ai (Fig. 1). Let's have a closer look at those vendors and their data mining and data analysis products and services.

Alteryx, Inc. (<https://www.alteryx.com>), founded in 1997, operates a self-service data analytics software platform. The main products/components are Alteryx Designer for data preparation and analytics; Alteryx Server, a secure and scalable server-based product for scheduling, sharing, and running analytic processes and applications in a Web-based environment; Alteryx Connect, a collaborative data exploration platform; Alteryx Promote, an analytics model management product for data scientists and analytics teams to build, manage, monitor, and deploy predictive models in production; and Alteryx Analytics Gallery, a cloud-based collaboration that allows users to share workflows in a centralized repository [1]. Fig. 2 shows a screenshot of the Alteryx Designer.

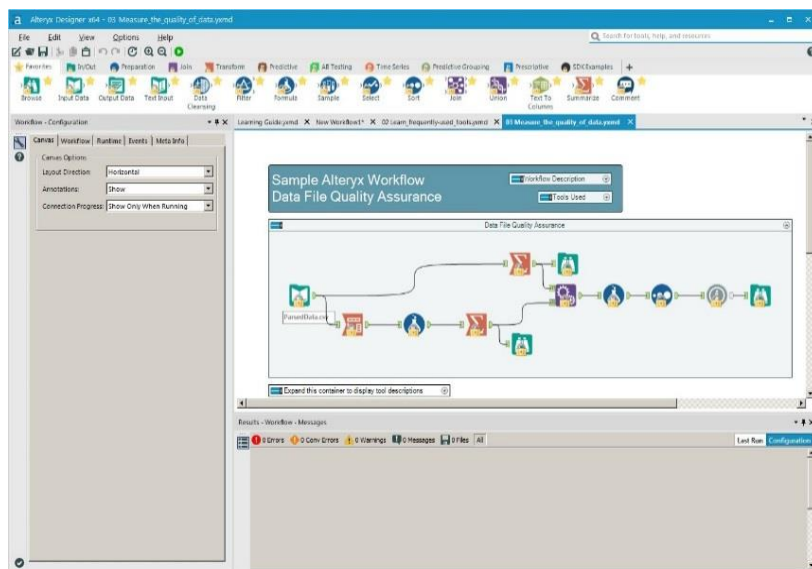


Fig. 2 Alteryx

SAS Institute (<https://www.sas.com>) provides analytics, business intelligence, and data management software and services.. SAS Institute started in late 1960th as a project at the North Carolina State University to create statistical analytics software (hence the name SAS). Since then, SAS Institute has become the largest privately owned software company in the world. The flagship product of the company is SAS, a huge system for advanced analytics, multivariate analyses, business intelligence, data management, and predictive analytics. SAS provides a graphical point-and-click user interface for non-technical users and more advanced options through the SAS language. Other products offered by SAS are JMP and Enterprise Miner. JMP is a suite of computer programs for statistical analysis developed by the JMP business unit of SAS Institute since 1989. SAS Enterprise Miner, as a stand-alone data mining tool, was released in 1999 (for a screenshot of Enterprise Miner version 7.1, see Fig. 3).

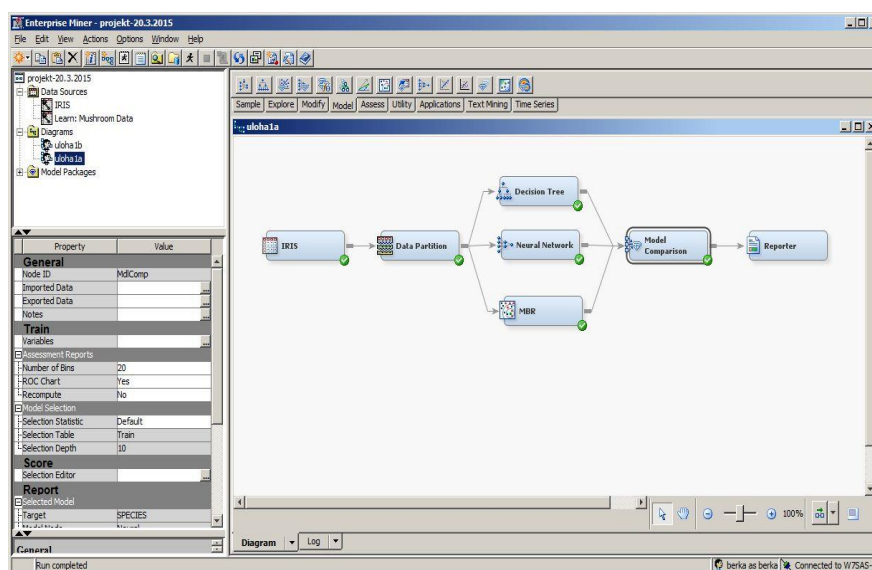


Fig. 3 SAS Enterprise Miner

RapidMiner, Inc. (<https://rapidminer.com>) develops an open source data science platform. The company focuses on the creation, delivery, and maintenance of predictive analytics. It offers RapidMiner Studio, a visual programming environment for predictive analytic workflows; RapidMiner Server that enables users to share, reuse, and operationalize the predictive models and results created in RapidMiner Studio; and RapidMiner Radoop that provides graphical environment for big data analytics using Hadoop and Spark.

The main data mining product is RapidMiner Studio, an integrated visual environment for data preparation, machine learning, deep learning, text mining, and predictive analytics (see Fig. 4. for a snapshot). RapidMiner Studio functionality can be further extended with additional plugins from the RapidMiner Marketplace in an easy way, similar to downloading applications into a mobile phone. The development of the system started in 2001 at the University of Dortmund, under the name YALE (Yet Another Learning System), so the company emerged as a spin-off at that university (the company is now based in Boston, USA). RapidMiner Inc. has a mixture business model. RapidMiner Studio is available both as a free system (the so-called community edition with the limit of 10 000 rows in the data table) and as a paid enterprise edition [1].

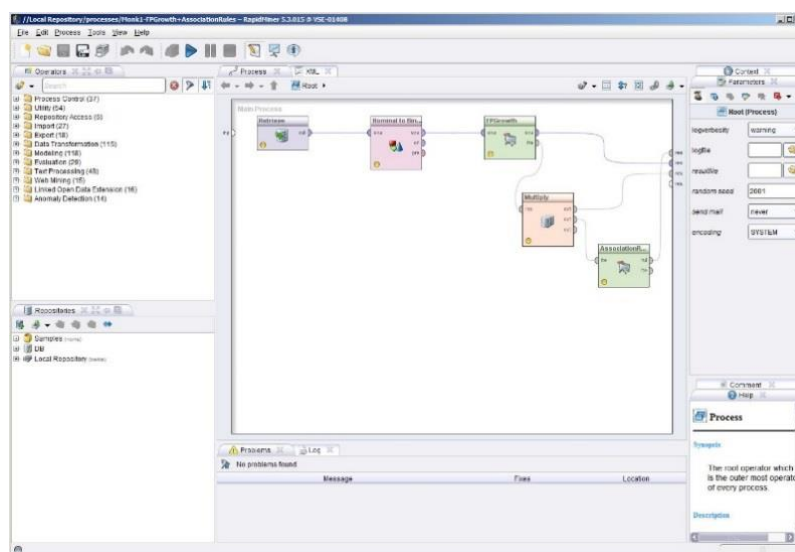


Fig. 4 Rapid Miner Studio

KNIME.com AG (<https://www.knime.com>), offering open-source enterprise solutions and services, is another vendor that started at a university - in this case, at the University of Konstanz, Germany, in 2004 (KNIME.com AG is now based in Zurich, Switzerland). The original system KNIME (KoNztanz Information MinEr) is a free and open-source data analytics, reporting and integration platform. Like RapidMiner, also KNIME offers various extensions to the system, now named KNIME Analytics Platform (see Fig. 5 for a snapshot). Beside this product, which runs as a desktop, KNIME.com AG also offers KNIME Server for remote and scheduled execution, automation, management, and deployment of data science workflows and tools for integrating KNIME Analytics Platform in large open source projects like Keras for deep learning, H2O for high performance machine learning, Apache Spark for big data processing, Python and R for scripting, and more.

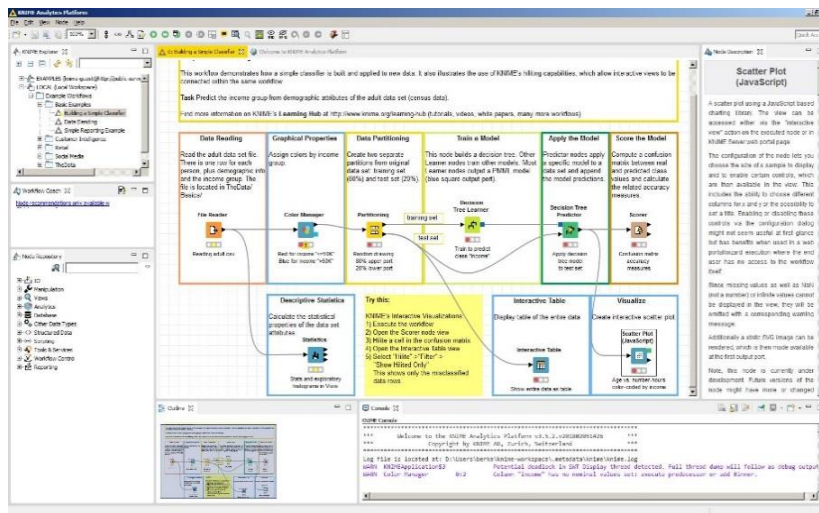


Fig. 5 KNIME

H2O.ai (<https://www.h2o.ai>), founded in 2011, is a relatively new vendor that appears in the Magic Quadrant in 2017 for the first time. H2O.ai develops a parallel processing prediction engine for machine learning and predictive analytics of big data. It offers H2O, an open source, distributed in-memory machine learning platform that supports the most widely used statistical and machine learning algorithms (see Fig. 6 for a screenshot) and H2O Driverless AI, an enterprise platform that employs the techniques of expert data scientists in an easy to use application, which empowers the users of the system to work on their projects faster using automation of different data mining steps. The company also provides marketing mix modeling, risk and fraud analysis, advertising technology, and customer intelligence solutions [1].

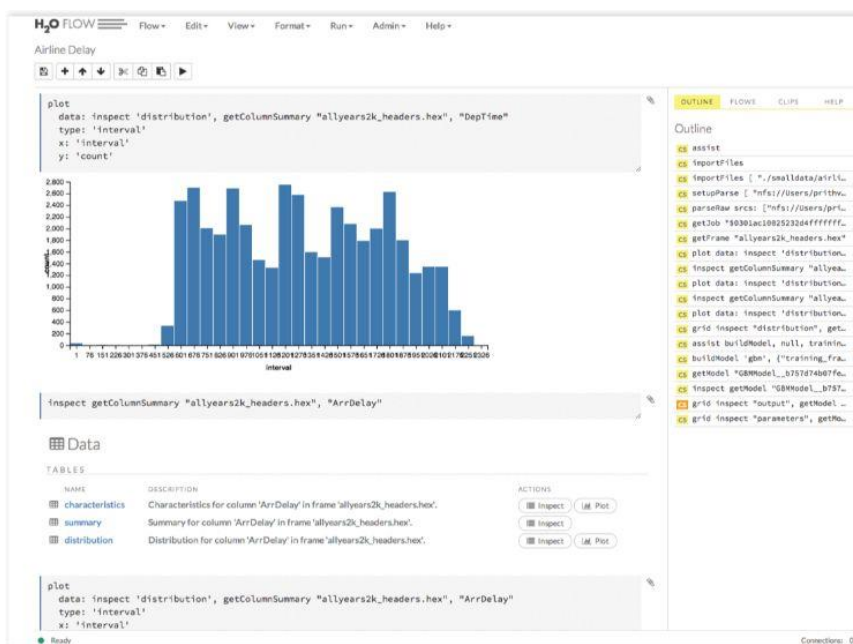


Fig. 6 H2O.ai

Not only the „snapshot“ of the vendors’ position in the Magic Quadrant but also the changes over time are interesting. Fig. 7 and Fig. 8 show the current leaders’ positions in the Magic Quadrant in the last five years (period 2014-2018). As we can see, all the leaders keep high completeness of vision during the whole period. Concerning the ability to execute, the established vendors also keep high ability to execute, while „newcomers“ improve this criterion. So we can conclude that the situation among the leaders is stable, as most of them keep their positions for several years.

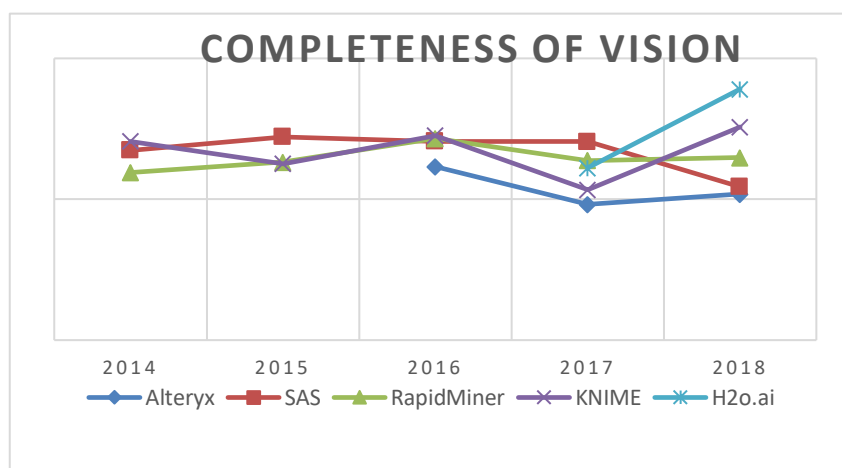


Fig. 7 Completeness of vision of MQ 2018 leaders in the last five years

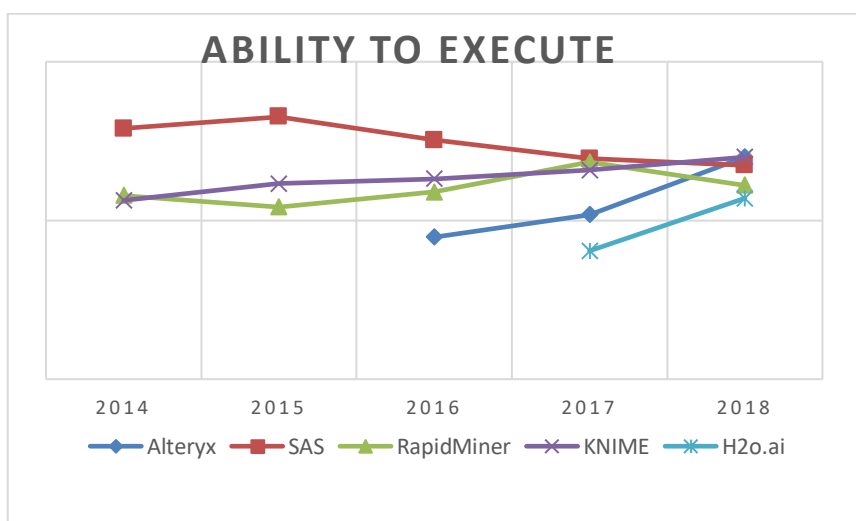


Fig. 8 Ability to execute of MQ 2018 leaders in the last five years

5 Conclusions

Magic Quadrant is a useful technique that allows to determine the strongest vendors on a particular market, for the purposes of this paper - on the market with tools suitable for data mining and data science. As the analysis of Gartner's Magic Quadrant report over the last five years shows, the situation among the leading vendors, as monitored by the Gartner, Inc., is stable. However, we must be aware that the Magic Quadrant focuses on commercial vendors. To get the full picture of the market with data mining and data science tools, we must consider also the open-source software community and, eventually, also the cloud solutions.

References

1. Internet Software and Services, Company overview, 2018. [online] Available at: < <https://www.bloomberg.com/research/stocks/private>> [Accessed 13 September 2018].
2. Gartner 2015 Magic Quadrant for Advanced Analytics Platforms: who gained and who lost, 2015. [online] Available at: < <https://www.kdnuggets.com/2015/02/gartner-2015-magic-quadrant-advanced-analytics-platforms.html>> [Accessed 20 July 2018].
3. Gartner 2016 Magic Quadrant for Advanced Analytics Platforms: gainers and losers, 2016. [online] Available at: < <https://www.kdnuggets.com/2016/02/gartner-2016-mq-analytics-platforms-gainers-losers.htm>> [Accessed 20 July 2018].
4. Gartner 2017 Magic Quadrant for Data Science Platforms: gainers and losers. 2017. [online] Available at: < <https://www.kdnuggets.com/2017/02/gartner-2017-mq-data-science-platforms-gainers-losers.html>> [Accessed 20 July 2018].
5. Gainers and Losers in Gartner 2018 Magic Quadrant for Data Science and Machine Learning Platforms, 2018. [online] Available at: <<https://www.kdnuggets.com/2018/02/gartner-2018-mq-data-science-machine-learning-changes.html>> [Accessed 20 July 2018].
6. How Markets and Vendors Are Evaluated in Gartner Magic Quadrants, 2017. [online] Available at: <<https://www.gartner.com/doc/3188318/markets-vendors-evaluated-gartner-magic>> [Accessed 20 July 2018].
7. SAS, IBM, RapidMiner, Knime leaders in Gartner MQ for Advanced Analytics Platforms, 2014. [online] Available at: < <https://www.kdnuggets.com/2014/02/gartner-mq-for-advanced-analytics-platforms.html>> [Accessed 20 July 2018].

Contact data:

Prof. Ing. Petr Berka, CSc.

University of Finance and Administration
Estonská 500, 10100 Praha 10, Czech Republic
berka@vse.cz