International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 1809–1823

Data Analytics Tools - A survey

Dr. S. J. Wagh [1], Kasturi Kulkarni [2]

Department of Information Technology
Institute of Government college of Engineering
Karad, Maharashtra, India
[1] Sanjeev.wagh@gcekarad.ac.in
[2] Kastukulkarni18@gmail.com

Abstract

The world is becoming more and more data-driven. Big data analytics is a rapidly growing research area in computer science and many other industries all over the world. The enormous amount of data has been rigorously generated and exchanged every day. Data analytics is a science of analyzing this data in order to make proper decisions and precise conclusions about the information. Data analytics tools are used in order to accomplish data analysis affairs in our modern businesses. Choosing the right data analytics tool is challenging, as no tool fits every need. In this paper, we will see a comprehensive list of data analytics tools that are obtainable and widely used in the market.

Keywords: *Data analytics, Data analytics tools, Big data*

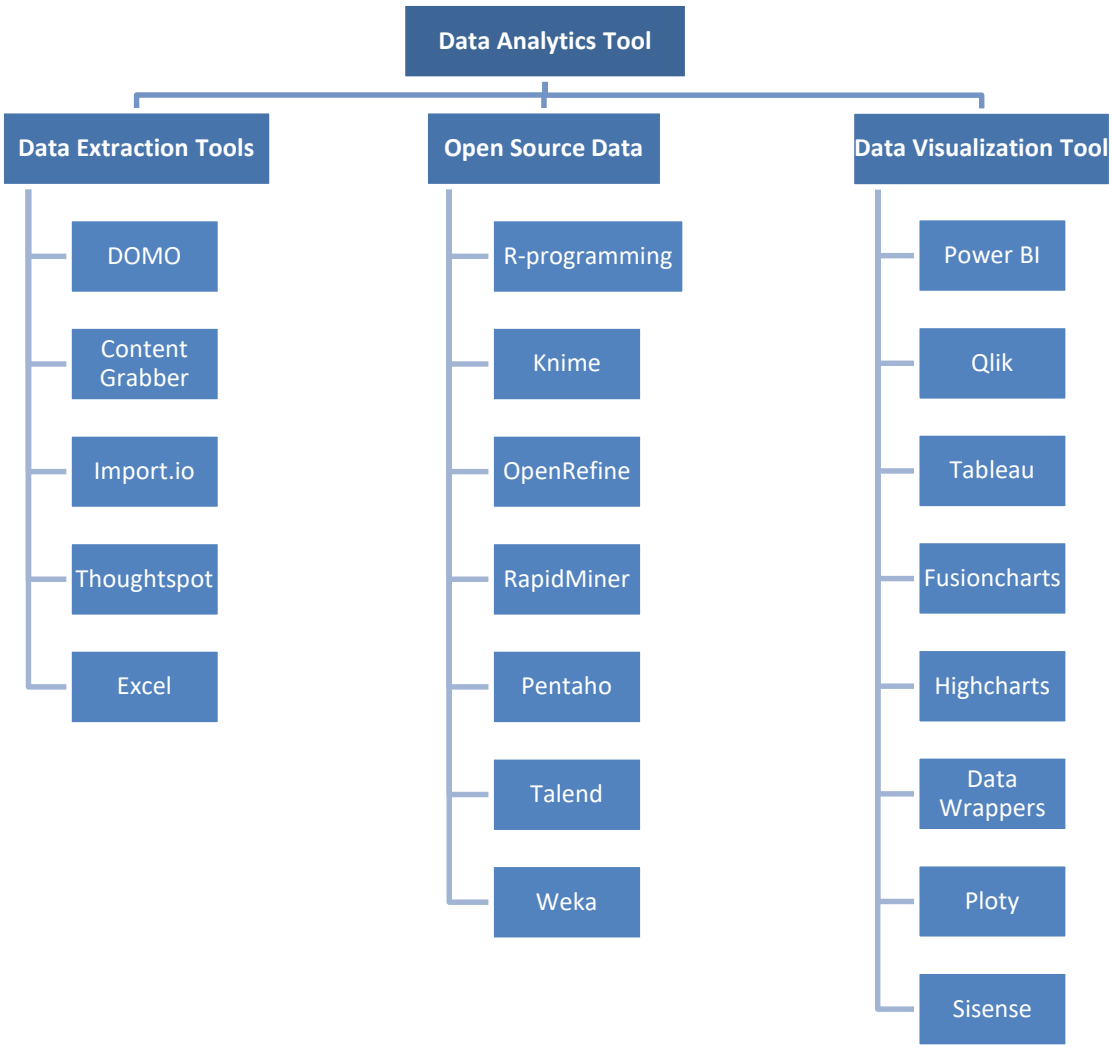
1. Introduction

The growing demand and importance of data analytics in the market have become preliminary offering worldwide. Before understanding the need of data analytics tools it is essential to understand the process of data analytics.
To make the whole process of knowledge discovery in databases (KDD) more clear, few operations have been included, which are pre-processing, transformation, data mining, and interpretation/evaluation. With these operators at hand, we will be able to build a complete data analytics system to gather data first and then find information from the data and display the knowledge to the user [1]. To summarize the process we first need to determine data requirement and data grouping. Secondly, data needs to be collected as data can be in any form and will be coming from a variety of sources. Once it is been collected it must be organized so that it can be analyzed. Then the data must be cleaned up before analysis. Now, the error-free data is analyzed. Each step in a process requires different tools for different purposes.

Big Data Analytics software is widely used in providing meaningful analysis of a large set of data. This software helps in many aspects such as finding current market trends, what are customer preferences, what's happening in industries, cutting operation costs, financial growth of company, forecasting business sales and other business information. It is also used by small businesses, retail companies, in medicine, and even in the world of sports [2].

ISSN: 2233-7857 IJFGCN

List and Classification of Data analytics Tools:



1. Data Extraction Tools:

1.1 DOMO

DOMO features many built-in data connectors and a visual data prep interface to accelerate data sourcing and transformation. Its robust business intelligence capabilities enable visualization and social commenting to facilitate collaboration. Domo also boasts native mobile device support with the same analysis, annotation, and collaboration experience as desktop.

From marketing to operations, HR to finance, IT to product development, supply chain to sales, see first-hand how Domo can change the way you do business and your ability to go fast, go big and go bold. It Serve the business with unbelievable speed, agility, and scale. Domo makes BI less work for IT and orders of magnitude faster for business. Solve back-end data integration challenges and put well-governed BI power in the hands of business users. It helps building apps on Domo's flexible, customizable platform [3].

With DOMO we can connect data from anywhere in the organization, Transform and normalize our data, Get immediate insights to raw data and Make better decisions using Domo's predictive analytics.

ISSN: 2233-7857 IJFGCN

1.2 Content Grabber

Content Graber is web crawling software for advanced extraction. It has a programming operation environment for development, testing, and production servers. You can use C# or VB.NET to debug or write scripts to control the crawler. It also allows you to add third-party extensions on top of your crawler. With comprehensive capabilities, Content Grabber is exceedingly powerful to users with basic technical knowledge.

1.3 Import.io

Import.io helps the world's largest companies strategize for success with smart web-data. Import.io is a web-based data extraction tool. Import.io removes the complexities of working with web data, allowing you to unify fragmented data from across the internet into something you can trust. By providing advanced capabilities like machine learning extraction, extensive content APIs, advanced data quality capabilities and so much more [4].

1.4 Thoughtspot

Thoughtspot features a search engine-like interface and AI to enable users to take a conversational approach to data exploration and analytics. Its SpotIQ engine parses search requests with which we can find insights hidden in their company data in seconds. Use search to analyse your data and get automated insights when you need them.

1.5 Excel

Excel is a basic, popular and widely used analytical tool almost in all industries. Excel becomes important when there is a requirement of analytics on the client's internal data. Excel has the advance business analytics option used in modelling capabilities.

The Analysis ToolPak is an Excel add-in program that provides data analysis tools for financial, statistical and engineering data analysis. The Analysis Toolpack includes different analysis operations such as Random Number generator, Fourier analysis, Descriptive Statistics, Exponential Smoothing etc. under analysis group.

2. Open Source Data Tools:

2.1 Apache Spark

Apache Spark is a unified analytics engine for large-scale data processing. In the recent years, Apache Spark has emerged as a solid foundation for data science and has taken the big data analytics domain by storm [5]. Spark is an in-memory cluster computing framework for processing and analysing large amounts of data. It provides a simple programming interface, which enables an application developer to easily use the CPU, memory, and storage resources across a cluster of servers for processing large datasets [6].

2.2 R-Programming

R-Programming is an excellent tool that can perform any sort of statistical analysis. It has gained credits and popularities in recent years due to its ease of use, extensive functionality and mainly for huge library packages. Data can be easily organised, implemented, manipulated, customized and re-organised with R programming [7].

2.3 Knime:

KNIME Analytics Platform is the open source software for creating data science. Intuitive, open, and continuously integrating new developments, KNIME makes understanding data and designing data science workflows and reusable components accessible to everyone [8]

A data science team builds a workflow in KNIME Analytics Platform which is based on an approach similar to collaborative filtering. It makes automatic predictions of individual customer interests by

ISSN: 2233-7857 IJFGCN

collecting preferences from many customers. The workflow starts with ETL and other data preparation steps, before creating a recommendation engine, and lastly through to determining interaction points for the Analytical Application. The workflow is then deployed on KNIME Server as a Guided Analytics Application [9].

2.4 OpenRefine

OpenRefine (previously Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data. OpenRefine always keeps your data private on your own computer until you want to share or collaborate. OpenRefine is available in more than 15 languages. It involves 1. Explore Data- n help you explore large data sets with ease. 2. Clean and Transform Data 3. Reconcile and Match Data – It can be used to link and extend your dataset with various web-services. Some services also allow OpenRefine to upload your cleaned data to a central database, such as Wikidata [10].

2.5 RapidMiner

RapidMiner provides a data science platform to help you drive real business impact. RapidMiner Studio is a powerful data mining tool for rapidly building predictive models. The all-in-one tool features hundreds of data preparation and machine learning algorithms to support all your data mining projects. It is a visual data science workflow designer accelerating the prototyping & validation of models [11].

2.6 Pentaho

Pentaho was created in 2004, comprising Pentaho Reporting, Pentaho Reporting Server, Mondrian OLAP Server and Pentaho Data Integration tools. These tools composed the Pentaho Open BI Suite. In 2006, Pentaho encompassed the Kettle and Weka projects. In 2015, Hitachi Data Systems acquired Pentaho. In the last years has been released a new edition per year, being currently in version 8.2. Pentaho highlights features that are only present in the commercial version. Pentaho Data Integration (PDI), as the ETL tool, extracts data from numerous sources, transforming and integrating it into the Enterprise Data Warehouse (EDW), either a Relational Database or NoSQL Database. Pentaho Data Integration can again be used for further transformation from EDW to Data Marts or Analytic Databases. Analytical features such as multidimensional analysis with OLAP or reporting can then be used alongside with other external BI tools [12].

2.7 Talend

Talend is a big data tool that simplifies and automates big data integration. Its graphical wizard generates native code. It also allows big data integration, master data management and checks data quality. It is open-source integration software designed to turn data into insights. It provides various services and software, including cloud storage, enterprise application integration, data management, etc.

It is a developer tool and job designer based on eclipse. To create and run ETL or ETL Jobs, Drag and Drop components and connect them. It automatically creates the Java code for the job instead of typing a single line of code. Talend Open Studio for Data Integration includes, Data migration from one database to another, Regular file exchanges between systems, and Data synchronization. It is the most inventive and effectual data integration solution on the today's world market. Business modelling, Real-time debugging and robust execution, Graphical development, Metadata-driven design and execution are the main features and benefits of Talend [13].

2.8 Weka

The full name of Weka is Waikato Environment for Knowledge Analysis. As a public work platform of data mining, Weka includes 10 java packages such as associations, classifiers, core, filters, etc. The package of Core is the core of the Weka system, including some key categories such as Attribute, Instances, etc [14]. Weka has more than 1 million lines of codes in total. The two packages of

ISSN: 2233-7857 IJFGCN

Associations and Core are the main java packages to achieve the association rules algorithm, which contain 190 java source files.

3. Data Visualization Tools:

3.1 Power BI

Power BI is a business analytics solution that will allow us to visualize our data and share insights across your organization, or embed them in your app or website. By connecting to hundreds of data sources and bring the data to life with live dashboards and reports [15]. Long story short, power BI make informed decision quickly, connect it, and model it, then explore your data with visual reports that you can collaborate, publish, and share.

3.2 Qlik

QlikTech is a software company founded in 1993 and of its products is the QlikView (QV) tool being a data discovery platform to support the decision making [16]. QlikView is Qlik's classic analytics solution for rapidly developing highly-interactive guided analytics applications and dashboards, delivering insight to solve business challenges. The modern analytics era truly began with the launch of QlikView and the game-changing Associative Engine it is built on. QlikView is a BI data discovery product for creating guided analytics applications and dashboards tailor-made for business challenges [17].

3.3 Tableau

From connection to collaboration, Tableau is the most powerful, secure and flexible end-to-end analytics platform for the data. It elevates every individual with the power of data. Tableau is the only business intelligence platform that turns your data into insights that drive action. Tableau combines a laser sharp focus on how people see and understand data with a robust, scalable platform required to run some of the world's largest organisation

Tableau Prep is how you get your data ready for analysis. Tableau's developer tools and APIs allow you to integrate, customize, automate, and extend Tableau to fit the specific needs of your organization [18].

3.4 Sisense

Sisense was founded in 2003 and is one of the prestigious BI tools for data visualization. Sisense supports data from several sources, allowing you to structure and import data sources but also to relate them and prepare them for analysis and visualization through ETL methodology (Extract, Transform, Load) [3]. Supported data sources can be files or data servers (SQL Server, MySQL, SnowFlake, etc.) located on local servers or web services like SalesForce and Google AdWords. Also have data connector alternatives, ODBC and REST (Representational State Transfer) [19]

3.5 Fushioncharts

Fusionchart mainly focuses on data visualization as data is meaningless without the ability to visualize, communicate and act on it. It is software that provides data visualization of a product or a process. The FusionCharts helps developers to convert any kind of data into meaningful and engaging visualizations, with the least amount of effort & learning on their part. The process involved is a fundamental human-technological activity that has the potential to help people solve important societal and scientific problems, and it can help people improve their understanding of complex phenomena through data [20].

3.6 Highcharts

Highcharts is a pure JavaScript based charting library meant to enhance web applications by adding interactive charting capability. Highcharts provides a wide variety of charts. For example, line charts, spline charts, area charts, bar charts, pie charts and so on. Charts are drawn using SVG in standard browsers like Chrome, Firefox, Safari, Internet Explorer (IE). In legacy IE 6, VML is used to draw the graphics.

ISSN: 2233-7857 IJFGCN

3.7 Data Wrappers

Data wrapper is a web-based tool that has enabled many people over the years to digitally build charts, maps and tables. There's an interface that a person can easily get to all of the charts once created. It is a four step process – (upload data, check and describe, visualize and publish and embed) to successfully publish our digital data [21].

3.8 Ploty

Ploty data visualization tool is helping leading organizations close the gap between Data Science teams and the rest of the organization. With Ploty, industries can easily design, develop, and operationalize data science initiatives that deliver real results [21].

Comparisons between different tools:

Tool	**Category**	**Platform**	**Features**	**Multipurpose Visualization**	**Advantages**	**Limitations**
DOMO	Data Extraction Tool	Cloud, Windows, Mac OS	1.Access is easy 2.Flexible business data connection options and integration 3.Project management 4. Social sharing tools. 5.Derive true insights into business data 6. Flexible deployment options	No	1. Integrates On-premise Data and External Data Sources in the Cloud 2. Visualizations are effective 3. Creates your dream dashboard 4. Optimized for mobile platforms	1. Expensive for most businesses 2. Lack of Improvement 3. Difficult to Extract Data
Content Grabber	Data Extraction Tool	Mac, Cloud, SaaS, Web. Installed - Windows	1.Customizable user interface 2. Agent editor and debugger 3. Data export and distribution 4. Agent logging and management tool	No	1.Performance 2. Scalability 3. Notification support 4. Scripting capability	1. Somewhat Inflexible 2. Unable to perform multiple scrapes at the same time
Import.io	Data Extraction Tool	Cross-platform	1. Automatically extract data from web pages into a	No	1.Fast, parallelized data acquisition	1.Does not support all scarping; additional

ISSN: 2233-7857 IJFGCN

			structured dataset 2. Authenticated extraction 3. Public APIs provide powerful and flexible access to extractors and crawl 4. Advanced Scheduling		distributed automatically by scalable cloud architecture 2.High availability for high volume usage 3.Flexible pricing options	scrapping features unavailable 2.Expensive
Thoughtspot	Data Extraction Tool	Windows, Mac, Linux.	1. Maximized Data Sources 2. Automated Insights 3. Unlimited Data Scaling 4. Cross-Source Analysis 5. Granular Access Control 6. Spot IQ insights 7. Single BI software	No	1. Expanded Cloud Offering 2. Easy usability 3. Search is very fast 4. Very intuitive design and the ability to share easily	1. Modelled for business owners needing to sort large amounts of data 2. Migrating between environments could be refined 3. Can do better for providing prescriptive analytics
Excel	Data Extraction Tool	Windows, Mac OS	1.Huge Capacity 2.Data Connectivity 3.Automated Reporting 4.Visualizations 5.Mobility 6.Collaboration, Distribution, and Publishing 7.Scalability and Performance 8.Security and Monitoring	No	1.Organizing Data 2.Streamlines Calculations 3.Multiple User Access 4.Third-Party Support	1.User Bias 2.Lack of Security 3.Time Consuming
Apache	Open	Microsoft	1.Fast	Yes	1. Dynamic	1. No

ISSN: 2233-7857 IJFGCN

Spark	Source Data Tool	Windows, macOS, Linux	processing 2. Flexibility 3.In-memory computing 4. Real-time processing 5. Compatibility with Hadoop 6. Better analytics		in Nature 2. Reusability 3. Support Multiple Languages 4. Cost Efficient 5. Support for sophisticated analytics 6. Active and Expanding Community	automatic optimizatio n process 2. Problem with Small File 3. No File Managemen t System 4. Less number of Algorithms 5. Manual Optimizatio n 6. Latency
R-Programmi ng	Open Source Data Tool	Cloud, Windows, Mac OS	1.Strong Graphical Capabilities 2.Highly Active Community 3.Wide Selection of Packages 4.Comprehensi ve Environment 5.Performs Complex Statistical Calculations 6.Distributed Computing 7.Data Wrangling 8.Compatible with Other Programming Languages 9.Vector Arithmetic 10.Data handling and storage	Yes-with plugins	1.R simplifies quality plotting and graphing 2.It is constantly evolving 3.R is predominant than others due to the development of statistical tools	1.Does not have support for dynamic or 3D graphics 2.Lesser speed 3.The steep learning curve makes it complicated 4.R lacks basic security 5.Sometime s working os slow
Knime	Open Source Data Tools	Windows, Mac, Linux	1.Big Data Extensions 2.Data Blending 3.Tool	No	1.easy connection of data sources, good handling of	1.Document ation is poor 2.Memory usage is

ISSN: 2233-7857 IJFGCN

			Blending 4.Meta Node Linking 5.Local Automation 6.Workflow Difference 7.Data Manipulation 8.Data Mining		the analytical model, easy to modify 2.Integrate different machine learning frameworks and techniques 3.Easy to isolate 4.Process large quantities of data 5.Text processing is easily performed by the various extensions	problematic some of the time 3.User interface can be crowded at times 4.Visualization can be improved 5.Sometimes working is slow
OpenRefine	Open Source Data Tool	Windows, Mac, Linux	1.Importing Data 2.Filtering/ Faceting 3.Editing cells/columns 4.Reconciliation 5.Exporting Data 6.Undo/ Redo	No	1. OpenRefine is a desktop application. It opens in the browser as a Local Webserver. So, the data is safe and it doesn't get uploaded to the Google server. 2. It has facets which are used to filter the data into subsets and these clusters can be customized and organised into meaningful data. 3. It has a Browser	1. The UI of OpenRefine is not user friendly. 2. Although the features and functions are strong, the UI make OpenRefine looks boring. 3. Besides, in the visualization, the function is not scalable. 4. For instance, OpenRefine give user a view of data, but the image is not big enough to figure out

ISSN: 2233-7857 IJFGCN

					based interface, and so can handle more data efficiently. 4. OpenRefine has a strong feature in extending data -- user can use it to find Meta Data and it can be used to correlate with it.	complex distribution. 5. Unfortunately Google has removed support for this tool, making few of its features redundant.
RapidMiner	Open Source Data Tool	Cross-platform	1.Data Partitioning and Replacement 2.Automation and Process Control 3.Bayesian Modeling 4.Clustering 5.Visual Workflow Designer 6.Modeling Evaluation	Yes	1.Robust features and user-friendly interface 2.Maximize the usage of data 3.It supports unified platform 4.Has broad connectivity	1.Sometimes there is processing speed issues 2.Not much user-friendly
Pentaho	Open Source Data Tool	Windows, Linux, Mac OS X	1.Data integration 2.Customizable Features 3.Intuitive dashboards 4.Performance Measurements	No	1.Powerful and thorough visualizations 2.Get real-time analysis of information through in-memory data caching 3.Exercise full control with interactive and customizable web-based drag-and-drop	1.performance issues 2. Troubleshooting results sometimes difficult.

ISSN: 2233-7857 IJFGCN

					dashboards 4.library full of filter functions 5.to blend information sourced from other pools of information	
Talend	Open Source Data Tool	Linux, Windows, Mac - supported	1.Shared Repository 2.remote deployment 3.Data quality and profiling with Data Cleansing 4.Customized ETL data extractions	Yes	1.Significantly reduced costs 2.Highly efficient 3.Compatibility with data source 4.Continuous integration reduces overhead of repository management 5.Provide lot of control for developers	1. problem with setting up Talend 2.Less community support 3.Not enough components or custom them to perform operations 4.Not greatly available and tough to run on Tier1 5.Sometimes slow in processing
Weka	Open Source Data Tool	Windows, Mac	1.Data pre-processing / classification 2.deep learning 3.Predictive Modeling 4.Data Regression and clustering 5.ML algorithm library 6.Data visualization	No	1.Portable 2.Easy to use 3.Adapted to create new ways to machine learning algorithms 4.Contains tools with multiple uses 5.Free online courses available	1.Limited literature available 2.Less documentation and online support 3.As it runs on java; it is memory intensive
Power BI	Data Visualization Tool	Microsoft Windows	1.Power BI uses Content Packs 2.Print/Customizable	Yes	1.create a rich personalized dashboard with	1.does not support records having a size greater

ISSN: 2233-7857 IJFGCN

			Dashboard 3.Custom Visualization 4. DAX Data Analysis Function 5.Natural Language Q&A 6.Datasets Filtration 7.Informative Reports		enterprise-level customization 2.can set dashboards for auto-refresh an d allow users to publish latest reports securely 3.Power BI comes with two variants, on-premise, and cloud offerings 4.inbuilt support for integration with MS excel 5.supports a broad range of backend databases 6.Power BI is inexpensive	than 250 MB 2.works with single dataset at one point 3.maximum size of dataset supports in Power BI is 1 GB 4.Crowded User Interface 5.Rigit formulas
Qlik	Data Visualizati on Tool	Microsoft Windows	1. Support Self-service visualization 2. Guided analytics apps and dashboard 3. Embedded analytics and reporting	Yes	1. Governed Creation 2. Managed Data Connections 3. Modern, Open API's 4.Manageabil ity 5. Security 6. Scalability	
Tableau	Data Visualizati on Tool	Windows, Mac, Web-based, Android, iOS X	1. Enterprise-ready, Enterprise-proven 2. Dynamic parameters 3. Viz animations 4. live data sources or data extraction from external data	Yes	1. Multiply data's potential 2. Secure 3. Flexible deployment 4.Powerful monitoring and management	1.High Cost 2. Inflexible Valuation 3.Poor After-Sales Support 4. Security Problems

ISSN: 2233-7857 IJFGCN

			sources as in-memory data			
Fushioncharts	Data Visualization Tool	Microsoft Windows, Linux, macOS		No	1.Active community 2.Supportive Helpdesk 3.customized for unique and specific implementations 4.Integration with most of the platforms is easier 5.Easily integrate with most of the platforms	1.Font and resizing are the areas of concern 2.Very expensive solution 3.Not very easy to set-up 4.The Interface is a little out-dated.
Highcharts	Data Visualization Tool	Windows, Mac, Android 2.0+	1.Open Dynamic API 2.Multiple Axes 3.Numerous Chart Types & Polar charts 4.Angular Gauges 5.External data loading	No	1.Customization can be done readily and easily 2.A variety of options for charts are available 3. Easy and flexible to use.	1. License is quite expensive. 2.nested config is a little confusing
Data wrappers	Data Visualization Tool	FreeBSD, Linux, macOS, OpenBSD, Windows	1.Live chart updates 2.Full chart styling customization 3.Own custom layout	No	1.No installation required for creating a chart 2. Adequate for beginners. 3.Available free of cost for use 4.Media Industry is the major end-user 5.Charts can be embedded in a short span of time	1.A complex chart such as Sankey cannot be built 2.Data security is an area of concern 3.The free version does not support customization, exporting or printing of

ISSN: 2233-7857 IJFGCN

						charts
Ploty	Data Visualization Tool	Windows, Mac, Linux	1.Multiple Deployment Options 2.Data White Labelling 3.Falcon SQL Client Support 4.Analytic Web Apps Development Support 5.Charts Import & Editing	No	1.Plotly is adaptable 2.software allows for statistical analysis 3.Easily import data to charts 4. Easy to use online editor 5.Active user base	1.difficult to visualize complex data or make complex, multi-field charts 2. There are limitations to the free version
Sisense	Data Visualization Tool	Windows, Mac with installed plugins, Linux	1.Data collection, filtering, consolidation & storage 2. Scalable data handling 3. Scalable data handling 4.covers the full range of business analytics 5.Single-stack architecture	No	1.Minimal Total Cost of Ownership 2.Sisense is robust, yet easy to use 3.Sisense analyzes data accurately and in real time 4. easily integrate Sisense with other products	1.The Elasticube ™ isn't that user-friendly 2.Sisense is a "heavy" application 3.Sisense's dashboards only interact on the web 4.navigation and filtering on the mobile platform could be improved

Apart from tools mentioned here, there are multiple other tools which are used worldwide. The technology doesn't stop there; it is growing rapidly with each single day. Having said that with tremendous demand of data analysis there are many new tools have been developed and launched in market with more advanced technology and features.

Global phenomena of using Big Data to gain business value and competitive advantage will only continue to grow as will the opportunities associated with it. Today almost every field of Researchers, scientists, business organizations, government agencies, advertising agencies, medical researchers use data analytics platform to carry out decision making task. Hence selection of right data within the larger data set to analyse the whole data should be appropriate.

11. CONCLUSION AND FUTURE CHALLENGES

This paper explores various Data analytics tools in terms of Big Data Process. It proposes a framework and detailed comparison for the selection of tool. As this survey attempt will help the researchers and professionals in finding right tool for the particular problem they may have, it is expected that future

ISSN: 2233-7857 IJFGCN

research may be carried out in analyzing industry-wise Data Analytics tool which is very much in need as there are numerous tools available. Even though the current data analytics tools are able to manage large amount of data, still many businesses are in a need of data handling and management. Undoubtedly, the available tools are evolving at rapid pace it creates a demand of more advanced tools.

REFERENCES

[1] Chun Wei Tsai, Chin Feng Lai, Han Chieh Chao and Athanasios V. Vasilakos 5* 2015, Big data analytics: a survey Tsai et al. Journal of Big Data (2015) 2:21 DOI 10.1186/s40537-015-0030-3.
[2] Athmaja S. Hanumanthappa M. Vasantha Kavitha. 2017. International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). A SURVEY OF MACHINE
[3] https://www.domo.com/why-domo
[4] https://www.import.io/developer-edition/
[5] Georgios Gousios. 2018. 2018 IEEE/ACM 40th International Conference on Software Engineering: Companion (ICSE-Companion), Gothenburg, Sweden 27 May-3 June 2018. Big Data Software Analytics with Apache Spark.
[6] Guller, Mohammed, 2015. https://www.springer.com/us/book/9781484209653, Big Data Analytics with Spark-A Practitioner's Guide to Using Spark for Large Scale Data Analysis
[7] Maddumage, C., & Dhanushika, M. P. (2018). R programming for Social Network Analysis - A Review. 2018 3rd International Conference on Information Technology Research (ICITR). doi:10.1109/icitr.2018.8736142
[8] https://www.knime.com/recommendation-engine-for-retailers
[9] https://hub.knime.com/knime/spaces/Examples/latest/50_Applications/43_Recomendation_Engine_for_Retailers/02_Guided_Analytics
[10] https://openrefine.org/ - A free, open source, powerful tool for working with messy data
[11] https://rapidminer.com
[12] Leite, N., Pedrosa, I., & Bernardino, J. (2019). Open Source Business Intelligence on a SME: A Case Study using Pentaho. 2019 14th Iberian Conference on Information Systems and Technologies (CISTI). doi:10.23919/cisti.2019.8760740.
[13] J, S., S, P., K, R., K, S., & G, N. (2019). Data Validation in ETL Using TALEND. 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS). doi:10.1109/icaccs.2019.8728420
[14] Shen, Y., Liu, J., & Shen, J. (2010). The Further Development of Weka Base on Positive and Negative Association Rules. 2010 International Conference on Intelligent Computation Technology and Automation. doi:10.1109/icicta.2010.676
[15] https://powerbi.microsoft.com/en-us/what-is-power-bi/
[16] André Lousa, Isabel Pedrosa, Jorge Bernardino, 2019. Evaluation and Analysis of Business Intelligence Data Visualization Tools. 2019 14th Iberian Conference on Information Systems and Technologies (CISTI) 19 – 22 June 2019, Coimbra, Portugal ISBN: 978-989-98434-9-3
[17] https://www.qlik.com/us- https://www.softwareadvice.com/bi/qlikview-profile/
[18] https://download.e-bookshelf.de/download/0007/6068/31/L-G-0007606831-0014536440.pdf - Arshad Khan- Jumpstart Tableau- A Step-By-Step Guide to Better Data Visualization
[19] Dwivedi, S., Kasliwal, P., & Soni, S. (2016). Comprehensive study of data analytics tools (RapidMiner, Weka, R tool, Knime). 2016. Symposium on Colossal Data Analysis and Networking (CDAN). doi:10.1109/cdan.2016.7570894
[20] Sahu, S. K., Jacintha, M. M., & Singh, A. P. (2017). Comparative study of tools for big data analytics: An analytical study. 2017 International Conference on Computing, Communication and Automation (ICCCA). doi:10.1109/ccaa.2017.8229827
[21] Sogodekar, M., Pandey, S., Tupkari, I., & Manekar, A. (2016). Big data analytics: hadoop and tools. 2016 IEEE Bombay Section Symposium (IBSS). doi:10.1109/ibss.2016.7940204

ISSN: 2233-7857 IJFGCN