Data Analytics tool

Abstract:

Data analytics tools can help deliver that value and bring that data to life. A lot of hard work goes into extracting and transforming data into a usable format, but once that's done, data analytics can provide users with greater insights into their customers, business, and industry.Data  analytics tools that allow you to analyze and model the data through visual programming. Developing A software tool that will use for understanding data and designing data science work flows and reusable components accessible to everyone.

Keyword:

Introduction:

The growing demand and importance of data analytics in the market have generated many openings worldwide. It becomes slightly tough to shortlist the top data analytics tools as the open source tools are more popular, user-friendly and performance oriented than the paid version. There are many open source tools which doesn’t require much/any coding and manages to deliver better results than paid versions e.g. – R programming in data mining and Tableau public, Python in data visualization.

Visual analytics (VA) make easier the bulk of complex information for better decisions. VA enables enterprises to understand data much more quickly and to make faster, better decisions. With sharp improvements in computing and data storage, it helps to businesses to solve relevant issues.Visualization is creating pictures in your mind as you read. You can think of it as making a movie of the writing. Visualization is important because it can help with our reading comprehension, make you feel more connected to the material, and create a more personal experience.Visual Analytics is the science of analytical reasoning supported by interactive visual interfaces. Today, data is produced at an incredible rate and the ability to collect and store the data is increasing at a faster rate than the ability to analyze it.

Architecture:

Literature Review:

Proposed system approach

## **Blend data from any source**

## ****Open and combine**** simple text formats (CSV, PDF, XLS, JSON, XML, etc), unstructured data types (images, documents, networks, molecules, etc), or time series data.

****Connect to a host**** of databases and data warehouses to integrate data from Oracle, Microsoft SQL, Apache Hive, and more. Load Avro, Parquet, or ORC files from HDFS, S3, or Azure.

****Access and retrieve data**** from sources such as Twitter, AWS S3, Google Sheets, and Azure.

## **Shape your data**

****Derive statistics****, including mean, quantiles, and standard deviation, or apply statistical tests to validate a hypothesis. Integrate dimensions reduction, correlation analysis, and more into your workflows.

****Aggregate, sort, filter, and join data****either on your local machine, in-database, or in distributed big data environments.

****Clean data****through normalisation, data type conversion, and missing value handling. Detect out of range values with outlier and anomaly detection algorithms.

****Extract and select features****(or construct new ones) to prepare your dataset for machine learning with genetic algorithms, random search or backward- and forward feature elimination. Manipulate text, apply formulas on numerical data, and apply rules to filter out or mark samples.

## **Leverage Machine Learning & AI**

****Build machine learning models**** for classification, regression, dimension reduction, or clustering, using advanced algorithms including deep learning, tree-based methods, and logistic regression.

****Optimize model performance**** with hyperparameter optimisation, boosting, bagging, stacking, or building complex ensembles.

****Validate models**** by applying performance metrics including Accuracy, R2, AUC, and ROC. Perform cross validation to guarantee model stability.

****Explain machine learning models**** with LIME, Shap/Shapley values. Understand model predictions with the interactive partial dependence/ICE plot.

****Make predictions**** using validated models directly, or with industry leading PMML, including on Apache Spark.

## **Discover and share insights**

****Visualize data****with classic (bar chart, scatter plot) as well as advanced charts (parallel coordinates, sunburst, network graph, heat map) and customise them to your needs.

****Display summary statistics****about columns in a KNIME table and filter out anything that's irrelevant.

****Export reports****as PDF, PowerPoint, or other formats for presenting results to stakeholders.

****Store processed data****or analytics results in many common file formats or databases.

## **Scale execution with demands**

****Build workflow prototypes**** to explore various analysis approaches. Inspect and save intermediate results to ensure fast feedback and efficient discovery of new, creative solutions.

****Scale workflow performance**** through in-memory streaming and multi-threaded data processing.

****Exercise the power of in-database processing**** or distributed computing on Apache Spark to further increase computation performance.

Conclusion

Reference: