

OpenText Magellan

A comprehensive AI & Analytics platform that augments data driven decision-making and unlocks insights from Big Data and Big Content.



End-to-end AI & Analytics platform with pre-built components



Drive user autonomy with self-service data prep, exploration, and analysis



Reduce the time, effort, and expertise needed to gather insights



Improve operations with insights from structured and unstructured sources

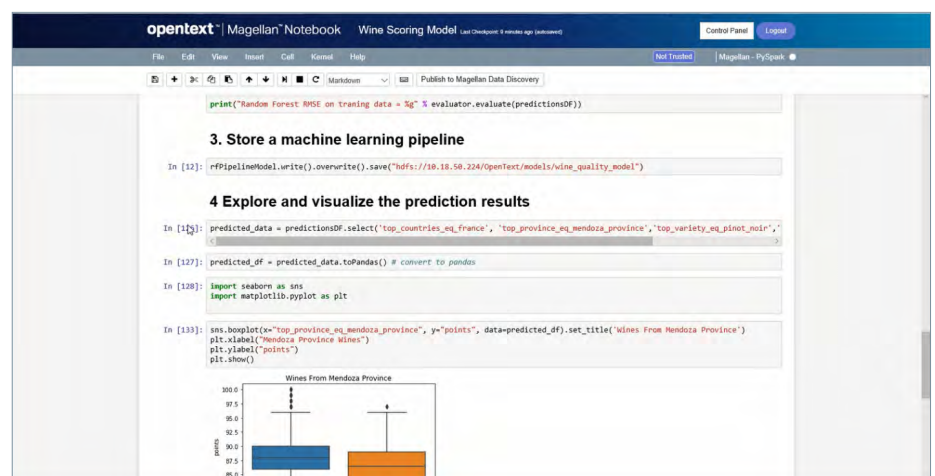
Associated OpenText products

- Magellan BI & Reporting
- Magellan Data Discovery
- Magellan Text Mining

OpenText Magellan is a flexible AI and Analytics platform that combines natural language processing and machine learning with advanced and predictive self-service analytics and business intelligence. Magellan enables business users to acquire, merge, manage, and analyze Big Data and Big Content from a wide variety of data sources. The flexible and intelligent Magellan is built leveraging a combination of open source and proprietary technology, with technologies like Apache Spark™ and Jupyter Notebook. This allows organizations to maintain ownership of their data, models, and algorithms, while taking advantage of a flexible, diverse, and high-performance platform. Magellan's cohesive, highly scalable infrastructure is perfectly equipped for handling massive amounts of structured and unstructured data, for numerous purposes.

Magellan Notebook

The Magellan Notebook provides data scientists with the ability to create custom algorithms and models to meet the unique and changing needs of their organization. Models and algorithms developed by data scientists can be published and then used by other Magellan components, where business users can leverage them to gather data insights without requiring deep technical knowledge. Each creation by a data scientist is saved to a data lake for fast and simple access within Magellan Data Discovery. The Magellan Notebook provides a familiar environment for technical users to help them get up and running quickly. It is built using the open source Jupyter notebook, and supports popular programming languages including Scala, Python, SQL, Java, and R, granting access to a broader technical talent pool. In notebooks themselves, data scientists can include explanatory text as notes or for easier collaboration, and visualizations can be created to vividly see results.



Data scientists can create visualizations within the Magellan Notebook and see it dynamically update as they write changes to it.

The Magellan Notebook also provides a set of capabilities to support machine learning (ML) model governance and notebook maintenance. It includes the ability to track changes made to notebooks through an integration with Git, allowing data

scientists to more easily collaborate with one another and trace the modifications made. The notebook also allows data scientists to maintain multiple versions of ML models, which can be used to improve model accuracy, test new models, and reduce the impact of data drift. It also provides an easy, out-of-the-box mechanism to validate ML models and determine if they have experienced degradation on a scheduled basis.

Category	Features
Methods to share results	<ul style="list-style-type: none"> • Create a machine learning pipeline for use in Data Discovery • Export results to HDFS. • Export results as a CSV file.
Built in support for additional python libraries	<ul style="list-style-type: none"> • Beautifulsoup • Nose • Numpy • Matplotlib • Pandas • Plotly • Scipy • Seaborn • Sklearn • Statsmodels • Yapf

Apache Spark

Magellan leverages Apache Spark™, the open-source general purpose computing engine, to deliver high performance, large scale data-processing and a powerful library stack across the AI & Analytics platform. Spark powers advanced analytics, ETL (Extract, Transform, Load), machine learning, data modeling, and enterprise BI capabilities within Magellan. It provides a consistent developer experience alongside the Magellan Notebook by supporting interactive use of popular programming languages, including Java, Python, R, and Scala, and offering over 80 high-level operators. Spark also unifies disparate functions within its processing framework in addition to its high-performance capabilities. It does so by providing multiple libraries out of the box that can be used seamlessly within Magellan:

- Spark SQL & Data Frames for data exploration and analysis
- MLlib for developing and publishing machine learning pipelines
- GraphX for graphs and graph-parallel computation
- Spark Streaming for building and accessing live data streams

A significant component to the Magellan Platform is the Magellan Data Discovery Gateway, which allows Magellan Data Discovery to natively connect to a supported data lake. The gateway provides a native connection directly between Magellan Data Discovery and Apache Spark on top of HDFS, which eliminates the need to transfer data from a data lake to another location for further analysis. Business users in Magellan Data Discovery can use the Magellan Data Discovery Gateway to leverage advanced algorithms and ML models, created and published by data scientists with the Magellan Data Science Notebook, to gain deeper levels of insight.

Additional Data Discovery capabilities available through Apache Spark

- **Aggregate** to aggregate or group values in a data field
- **Decode** to change the name to something more recognizable, such as Male and Female
- **Numeric Ranges** to create a unique field in which the values from a numeric field appear as a set of numeric ranges that you define
- **Quantile** to create a group from a numeric field that contain the same quantity of values
- **Ranking** to rank a table based on a specific column
- **Expressions** to create a calculated column from one or more database columns

Magellan Data Discovery

OpenText™ Magellan™ Data Discovery allows business users and analysts to perform powerful advanced and predictive analysis on billions of records through an intuitive user interface. It also includes Smart Data Discovery, which recommends visualizations that are accurate and representative for the data selected, reducing the time and expertise needed to gather data insights. Techniques such as regression analysis, clustering, and crosstabs can be leveraged without requiring any coding knowledge alongside custom algorithms or models that are developed and shared by data scientists in the Magellan Notebook. Users can easily toggle between Magellan’s purpose-built analytical repository or enterprise Spark data lakes from within Magellan Data Discovery, depending on which set of powerful capabilities they need. The analytical repository provides a high-speed database with advanced data mining features such as correlations, clustering, forecasting, classifications and regressions. Apache Spark, meanwhile, provides access to custom algorithms and ML models published by an organization’s data scientists, and capabilities for high-performance analysis across vast datasets.

Category	Features
Connectivity	<ul style="list-style-type: none"> • Relational databases via pre-installed ODBC drivers or native driver installation • Cloud-based data source – Remote Data Discovery server • Network data sources – FTP & HTTP • Flat file formats – CSV, fixed-length flat files, Microsoft® Excel®
Export methods	Magellan BI Formats (BIRT .rptdesign & .datadesign), FastDB table, PDF, CSV, export to Spark/Hadoop
Types of data analysis	Clustering, set analysis, classification analysis, trend analysis, long tail analysis, time series analysis
Data transformation commands	15+ including create link, expression, parametric, aggregate, decode, parametric and ranking
Out of the box visualizations	15+ types of visualizations, charts, and indicators including dial, meter, bar, pie, scatter and tree map

[Learn more about Magellan Data Discovery](#)

Magellan BI & Reporting

Magellan BI & Reporting is a scalable analytics and data visualization platform that allows non-technical users to create, share, and consume engaging reports, dashboards, and visualizations from numerous data sources through a drag and drop interface, and intuitively dive into creations through click interactivity. Visualizations can be embedded into applications and web locations to place them in context and simplify the end user experience, while maintaining all interactive and data exploration capabilities. Magellan BI & Reporting also includes full-fledged, out-of-the-box ETL capabilities by seamlessly embedding Apache NiFi, one of the most capable technologies for enabling the automation of dataflows between virtually any set of data sources. This feature, called Magellan Dataflow, provides a flexible and reliable mechanism to move the data to where it is needed and prepare it for reporting, analysis, machine learning, archiving or other purposes.

Magellan Text Mining includes six modules that work together to extract, analyze, and process semantic metadata.

- **Concept Extraction** to identify meaningful keywords, phrases, and core concepts
- **Named Entity Recognition** to locate and identify terms or sequence of terms with a name, including different names for the same entity
- **Text Classification** to index and sort documents by classification, weight, and relevancy
- **Text Summarization** to identify key sentences and create an overview from topics of interest
- **Sentiment and emotion analysis** to determine the subjectivity, tonality, and emotion of content
- **Language Detection** to recognize languages and categorize documents based upon those

Category	Features
Out of the box connectivity	30+ data sources, including Oracle, Hadoop, HBase, and Cassandra
Visualization types	15+, including heatmap, scatter, and multi-series
Localized user interface languages	Chinese, English, French, German, Japanese, Spanish

Magellan BI & Reporting is comprised of four components that allow for users to interact with the platform based upon their needs and level of technical knowledge

- Magellan Interactive Viewer – Explore BI & Reporting reports and dashboards through click interactivity
- Magellan Dashboards – Aggregate multiple data sources into a visualization that meets for frequent and cadenced needs
- Magellan Analytics Studio – Self-service ad-hoc reporting tool for unique visualization needs
- Magellan Analytics Designer – Powerful and advanced IDE for creating dynamic reports and applications

[Learn more about Magellan BI & Reporting](#)

Magellan Text Mining

OpenText Magellan Text Mining allows organizations to tap into the insights held within their unstructured data, from external social media content to internal content, like documents, and emails. Instead of manually tagging metadata from unstructured data, organizations can automate such repetitive tasks with great scale and efficiency, while also improving content findability. Semantic metadata gathered through Magellan Text Mining can be used in-turn with other components of the Magellan Platform to provide additional sources of insight to make data-driven decisions

Category	Features
Connectors	<ul style="list-style-type: none"> • OpenText sources – Archive Center, Content Server, Documentum, eDocs, InfoArchive, Media Manager, and Tempo Box • 3rd Party repositories – Box.com, CMIS, DropBox, Google Drive, Google Gmail, IBM FileNet, Microsoft Exchange, Microsoft SharePoint
Methodology and analysis	<ul style="list-style-type: none"> • Pattern Matching, both statistical and linguistic • Machine Learning – Application of families of algorithms for ML models, and post-processing algorithms • Knowledge Engineering for taxonomies and authority files, and to improve ML prediction accuracy

[Learn more about Magellan Text Mining](#)

About OpenText

OpenText, The Information Company, enables organizations to gain insight through market leading information management solutions, on-premises or in the cloud. For more information about OpenText (NASDAQ: OTEX, TSX: OTEX) visit: opentext.com.

Connect with us:

- [OpenText CEO Mark Barrenechea's blog](#)
- [Twitter](#) | [LinkedIn](#)

 Join the conversation

 Keep up to date

 Learn more