



An automated predictive datamining tool

Annexe 3: TIMi Technical Features v1.06

August 2018

TIMi

Minimum system Requirements for installation on a server or workstation: see annexe 4: Technical requirements

TIMi is a suite of tool that includes the softwares: TIM Modeler, StarDust, Anatella and Kibella.

The « TIM Modeler » software allows you to create very accurate predictive models for any kind of prediction: binary targets, continuous targets, multi-class targets. The predictive models built with TIM are always amongst the “top winners” to all the international datamining competitions (PAKDD2006, KDD2009, AUSDM2009, PAKDD2010,...) in which TIMi sprl participate regularly. These competitions demonstrate the superior accuracy of the predictive models delivered by TIM. For more information at this subject: <http://www.business-insight.com/forum/viewforum.php?f=12>

The « StarDust » software allows you to explore and visualize in real-time 3D, in a multivariate way, your databases. It also computes in real-time segmentation models. It allows you to easily explain from a “business” perspective each of your segments. To use StarDust, you need a 3D-hardware-accelerated-graphical card that supports OpenGL1.5 (any recent computer will do the trick).

The « Anatella » software is an ETL tool built especially for the analytical needs of the modern enterprises. Anatella contains some unique functionalities tailored to improve and to ease the modeling process (automatic creation of derivative variables, integrated debugger, etc.). As all ETL tool, Anatella allows you to gather&process all your data in order to create datasets to « inject » into TIMi modeler, Stardust or Kibella. Anatella is the most user-friendly and the most versatile ETL tool available on the market (and one of the fastest).

The « Kibella » software is an unlimited self-service dashboarding tool for “classical” BI. Why “unlimited”? Because Kibella has no limits on the number of designers, viewers, dashboards, servers. Kibella runs inside the famous “Apache” webserver so that you can deploy it everywhere without fear of a hacker destroying your server.

Optional Component: The “TIMi Hadoop Module” is an Anatella extension that allows you to read & write “Hadoop Parquet Files” (This is the most common storage used inside Hadoop to store any table: e.g. These Parquet files are used by Spark, Hive, etc. to store all their data). The Anatella Parquet File reader&writer is made using optimized C code that is 2 to 20 faster than the equivalent standard Hadoop code. This means that manipulating Hadoop tables using Anatella is much faster than using any other tool in the Hadoop ecosystem (including Spark). This additional component also includes a direct access to HDFS to Read&Write any supported files (.parquet, .gel_anatella, .txt, .json, .xml, etc.) directly to/from any HDFS drive.

Optional Component: The Anatella “Privacy Module” is an extension for Anatella that adds the possibility to completely hide the content of the columns containing private/confidential informations. You can however still: compute aggregate on these columns, create predictive models on these columns, or do any type of computation on these columns (but without seeing them at the row level). Any data manipulation on confidential columns will only be authorized for users that are duly authenticated (through Kerberos/Microsoft Active Directory) so that it’s impossible to use/see confidential columns outside your private network. In this way, even if a thief is leaving with some data files, he won’t be able to use/see them (because he won’t have the proper authentication). A dedicated interface is provided to the CDPO (Chief Data Privacy Officer) to decide which columns are confidential. A network logger logs all accesses to confidential data. *This update makes Anatella the only data transformation engine built to safely handle the GDPR requirements. Using any other data transformation engine exposes you to the heavy financial sanctions of the new GDPR law in case of confidentiality breach (i.e. 4% of global turn-over).*

Your license to TIMi grants you access to the following functionalities (non-limitative list):

- TIMi Modeler: ELD functionality: Extremely Large Datasets (unlimited rows & columns).
- TIMi Modeler: extremely fast and advanced analytical engine that creates accurate predictive models for Binary prediction (will the prospect buy or not?), continuous prediction (how much will the prospect buy?) and multi-class prediction (which product will the prospect buy?).
- TIMi Modeler: The analytical engine of TIMi (which creates the predictive models) is 100% automated. It is not necessary loose time to refine some incomprehensible meta-parameters of the software to obtain (very) good models because the default settings are already optimal in 99% of the cases. However, TIMi sprl provides to the advanced users of TIMi with the opportunity of modifying all the parameters of all the algorithms used in the modeling process for a perfect control and a more complete understanding of the modeling mechanism.
- TIMi Modeler: Complete documentation of the algorithms. TIMi is not a "black box". We have a transparency policy towards our customers
- TIMi Modeler: Full documentation of the XML configurations files of TIMi . To help you to integrate TIMi into your production line in the easiest way.
- TIMi Modeler: All computations can be conducted in "interactive mode" (with the mouse) or in "batch mode" (without any user interaction) to enable the complete automation of the creation / use / update of the predictive models.
- TIMi Modeler: There are two computing modes to create a predictive model:
 - **"Insight"**: This mode generates the simplest models (usually less than 15 variables) and these models are usually already extremely accurate.
 - **"Performance"**: this mode generates more complex models (up to 500 variables for text-mining applications, for example) without sacrificing the accuracy of the model.
- TIMi Modeler: Estimation of the quality of the generated predictive models based on many Lift charts. Estimation of the lift stability: a confidence interval of the lift is directly generated.
- TIMi Modeler: high-performance module for creating predictive models to identify potential frauders (fraud detection) or potential churners.
- TIMi Modeler: Data Audit (simple OLAP reporting fonctionnality).
- TIMi Modeler: Possibility to define "On-the-fly" the target to predict, using a "business" expression (similar to an SQL query).
- TIMi Modeler: Possibility to define "On-the-fly" the segment of the population to be used for the modeling. This segment can be a "business" segment (selection of individuals using an expression similar to SQL) or an analytic segment computed by "Stardust".
- TIMi Modeler: Extraction of the profile of your target (customers/frauders) from the modeling results. This profile is easily understandable by a un-trained "business" user.
- TIMi Modeler: The target's profile is presented under the form of MSWord and MExcel reports containing many simple, pertinent and intuitive charts that accurately characterize your target.
- TIMi : Simple data manipulation tool: format conversion, splitting & sampling of datasets (used mainly for benchmarking).
- Anatella : ETL: advanced data manipulation tool:
 - Join tables,
 - Columns & Rows filtering out of tables,
 - Sorting,
 - Format conversions (CSV, SQL ,...),
- Derivation of new columns for modeling:
 - Automatic generation of hundreds of thousands of new, derivate columns ;
 - a full scripting language based on JavaScript (standard ECMA-262) that allows you to express the most complex transformations, validations, aggregations, derivations. This language is similar (but more versatile) to a "SAS datastep"
- Anatella: the development environment to create the data manipulation scripts is extremely simple, intuitive & versatile. This environment is based on a hybrid technology:
 - The simple transformations are described using "little boxes" (which is the most intuitive way to represent a data transformation and is a "de facto" standard for all modern ETL).
 - Complex transformations are programmed using a scripting language based on JavaScript (standard ECMA-262) which is simple, complete and very versatile.
- Anatella : Direct access to an integrated "debugger" with an interface similar to the Visual Studio debugger (to "debug" the scripts written in JavaScript/ECMA-262): you can add "break points" to your code, add some "watch" on variables, see the "stack",...
- Anatella : The data transformation operators are "**meta-data free**" which means it is not necessary to define "metas-datas" to use 99% of the various transformations available in Anatella. This feature is very important because, in datamining, it is common to manipulate tables of tens of thousands of columns and it is impossible to specify "by hand" the "meta-datas" of all these columns (...and that exactly what's all the other ETL tools force you to do).
- Anatella: Data transformations to apply BEFORE predictive analytics:
 - **Social Network Analysis (SNA)**: This plugin extracts some valuable social-metrics out of graphs Typically: The best connected individuals, the individual who plays the most important role in any group (i.e. the social leaders), etc. These social-metrics can thereafter be used in combination with a predictive analytic tool like TIMi to create very accurate Cross-selling, Up-Selling or churn marketing campaigns.
 - **Textmining** : The textmining-operators apply the classical "bag-of-word" technique to produce, starting from raw, unstructured text-data, many new columns and new variables directly exploitable inside TIMi or Stardust, for predictive analysis. You can now exploit the information contained inside any unstructured text-data (for example: text comments on your website, news articles, ...) to make any predictive model (for example for cross-selling, upselling and churn)!
 - **Data cleaning**: text-spelling correction: Anatella includes a unique operator that checks & corrects the spelling mistakes in any text field (clustering of similar words together).
- Anatella: Data transformations to apply AFTER predictive analytics:
 - **Apply your TIMi predictive models**: Anatella can natively apply TIMi predictive models. This allows you to easily score your database, along with other data transformations....
 - **Apply your Stardust segmentation models**: Anatella can natively apply Stardust segmentation models. This allows you to easily segment your database, along with other data transformations....

- Operational Research (OR) optimization toolbox: Anatella integrates:
- an efficient **multi-threaded LP/IP solver** that allows you to solve large scale optimization problem. The LP solver handles millions of constraints and several thousands of variables.
- an efficient **solver for the GAP (General Assignment Problem)**. Typical GAP problems includes: Which product do I have to offer to which customer when I have the following business-constraints: The stock of each product is limited, Each selected customer receives a folder with N offers (no less and no more than N offers), The margin on each product is different,... The GAP solver included inside TIMi handles campaigns with several millions of customers and thousands of products.
- The Anatella Optimization plugin is typically used for operation research (OR), sales & profit optimization, stock optimization, etc.
- Anatella: OLAP reporting through the usage of a "Microsoft Office Data Injection operator": This plugin allows you to automatically update any chart or graphics contained in any Microsoft Office document using some data extracted from any database(s) or flat file.
- Anatella: ability to read in natively compressed raw CDR files (i.e. Call Data Record files compressed with RAR, ZIP or GZ) (for telecom). For example: Anatella can extract more than 970 features out of raw CDR files create with an Ericsson 100313-1505. This is tremendous opportunities for a dataminer!
- Anatella: export any dataset to native Tableau file format (.tde), to native QlickView file format (.qvx)
- Anatella: reads natively .dbf;.dbfx;.xbf files (dBaseIII, dBaseIV, dBaseV, dBase level7, Visual dBase, Paradox, Foxpro, Visual FoxPro, Shapefile, Clipper DBF and DBT).
- TIMi, Stardust, Anatella: ability to read in "native" mode dataset files in text format (CSV) compressed RAR, ZIP, GZ.
- TIMi, Stardust, Anatella: direct native access to SAS datasets (.sas7bdat files)
- TIMi, Stardust, Anatella: Direct access to all relational databases via classical ODBC connections & OLEDB connections (Teradata, Oracle, SqlServer, MySQL, ...)
- TIMi Modeler: "Predictive Models" are saved as XML files. You can apply these predictive Models to any database an unlimited number of times.
- TIMi Modeler: Exportation of predictive models to:
 - Base SAS script (in the form of a "data step")
 - SQL (SQL script standard working in all DB's)
 - PMML,
 - C
 - HTML / JavaScript
 ... for a simple and automatic scoring of all your prospects. Using your models has never been easier!
- TIMi Modeler: "Model Merger" module: automatically chooses the right models to apply to your database. The selection of the

- models is based on the different segments that you have defined. These segments can be "business" segments (selection of individuals similar to SQL) or segments calculated by "Stardust". Ability to apply simultaneously 1 thousand different models on the same dataset for Cross-selling purposes.
- TIMi Modeler: Automatic analysis of the degree of expiration of the predictive model (Is the predictive model not too old?), Possibility to replace the old model by a new model re-computed in real time.
 - TIMi Modeler: Analysis of over- or under-commercial exploitation of a geographical area/channels/categories/ etc. or groups of clients to enable the optimization of investment in human and advertising for each segment (geographic analysis is possible if location data as the "zip/postal code" is available for your customers).
 - TIMi Modeler: Real time module that simulates the "ROI" (return on investment) of a marketing campaign based on a predictive model built using TIMi . This module allows you to (amongst other things) optimize the size of your marketing campaign.
 - Stardust: Exploration and Visualization in real-time 3D in a multivariate way of your entire database (possibility to visualize in real-time and in a "fluid" way a database of several million of individuals).
 - Stardust: Detection of (multivariate) "Outliers" and non-valid data.
 - Stardust: Creation of segments based on your entire database using segmentation algorithms familiar to all data miners, such as:
 - PCA algorithm,
 - quantile-recoding algorithm,
 - "Minimum spanning tree" algorithm,
 - many variations of the ward's algorithm,
 - many variations of the "K-Means Algorithm", etc.
 You can also use many different variations of the "distance definitions" between the individuals: Pearson-Distance, Distance, Cosine, Euclidean distance, etc.
 - Stardust & TIMi Modeler: 3D visualization of the "target" population of the predictive model and multivariate characterization of this target. Analysis of the different segments that compose the target.
 - Stardust: Automatic creation of an analytical report in HTML format that describes from a "business perspective" each segment. This report contains many pertinent and intuitive chart for the "business" user.
 - Kibella: Unlimited self-service dashboarding.

The installation of TIMi on a machine is made using a conventional and user friendly setup program. You must have administrator rights to install TIMi. The installation usually takes less than 5 minutes.