

DEFINITION OF REQUIREMENTS

DEMAND FORECAST

SMART FACTORY 4.0

Contents

Version Control.....	3
General Requirements	4
Input Requirements.....	4
Model Requirements	12
Performance Evaluation Requirements	13

Version Control

Delivery date	Version	Description	Modified chapters	User
08/05/2020	0.1	Business requirements – Sprint 1 deliverable	Business requirements	Closer

Annexes

Annex	Name	User	Delivery date
1	Industry4.0.1.docx	Closer	24/04/2020

General Requirements

The requirements, or rules, under which the demand forecasting module will be developed are:

- The forecasting module is to be implemented on a remote Windows or Linux server that connects via API both to local databases or local file servers to access the client's raw input data and to external sources as described in detail below. Outputs are made available through online dashboards and are accessible through a custom API.
- An information pipeline, transferring the demand forecasting module's outputs as inputs to the PPO module will also be considered.
- This system will consider a general implementation, applicable to any area of industry, with multiple customisable features that can be activated or deactivated by the client in an initial interface. These features are:
 - Input variables to consider;
 - Unavailable inputs that should be estimated from the data;
 - Time frame of the forecast;
 - Model(s) to use;
 - Predefined evaluation metrics to display;
 - Choice to see final decision dashboard (to perform manual configuration of the models) or not, alongside the demand predictions.
- Input data loading procedure will be automated as much as possible, while maintaining the possibility of manual parametrization by users. A pre-processing pipeline will be implemented to format the input data as required by each of the three models.
- On the prototype, demand forecasting will be performed at single factory level for each single product or product category. For each factory, the products will be categorized based on orders frequency. Additional customisation options can be provided to determine granularity in subsequent phases (if possible).
- All time frames considered in the models will be customisable in the initial screen, to allow for different forecasting schedules, in relation to the time period most relevant to the client's business. The default setting for the prototype will be a monthly forecast, with the lowest available granularity level of the raw data as a limit. Finer granularity analyses may be provided as an additional service, pending consultation.

Input Requirements

Although this module strives to offer as many customisation options as possible, the data pre-processing pipeline requires the observance of certain rules in relation to the format of the raw data provided by the client.

- Major requirements for Inputs processing are the following: Input data will potentially be loaded from 3 major sources:
 - Company's Internal Systems. Which can imply a direct connector to a data warehouse, if it is in place (SQL, NoSQL, Spark, etc.), or a loading pipeline if data stored directly in Files (i.e. CSV, Excel etc.), or a mix of the two.
 - External pre-identified web sources, through custom APIs and/or scrapping scripts.
 - User interface, with manually controlled inputs in this case.
- The very minimum input data must include at least the first 4 lines of Table 1 (mandatory variables).
- If the client had any previous forecast system, the tool will require that data to generate some possible information to our model. In this case, the format must mirror the one of demand data, as shown in Table 1.
- Exogenous variables vary in importance depending on business activity. The configuration parameters of the forecast module will allow the selection of the set of exogenous variables that should be considered to generate predictions. Within each set, it will be possible to select each single variable separately. Default option, in case no variables are specified, is to use the entire exogenous set.
- Likewise, option will be given in the configuration parameters to choose between internal and pre-identified external data sources, for all those exogenous variables where the choice is possible.
- If data related to marketing event/promotion dates is unavailable, an option is given to use estimated values based on registered peaks in the number of orders.
- Likewise, if data related to macroeconomic events that negatively impacted the business is unavailable, an option is given to use estimated values based on registered lows in the number of orders.

Table 1 – Summary of considered inputs

Variable Name	Importance ¹	Description	Input Format	Source	Loading	Comments/Roles
Order Date	MH	Univocal tag for reception date of the order	Datetime object "DD-MM-YYYY"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Time reference for ALL Timeseries of data (Index)
Stock Keeping Unit (SKU)	MH	Univocal tag for single product Group products by Category based on orders frequency	String "ABC012#"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Filter Signal to perform forecasts by product; and/or group of products

¹ MH = Must Have; C = Critical; I = Important; BD = Business Dependent; O = Optional

Client ID	MH	Univocal code/name of client for each order	Int OR String "ABC012#" --> Bin "0;1;2"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Additional KEY to merge signal and exogenous data in pre-processing phase. IF KEPT: Grouped and encoded ELSE: Dropped
Ordered Quantity	MH	Quantity of products ordered in uniform metric	Int/Float	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	SIGNAL to be forecasted over different time horizons
Ordered Units	C	Lots of products ordered	Int	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	IF Ordered Quantity Normalized: alternative Signal to be forecasted; ELSE: Dropped
Client Previous Prediction						
Order Date	MH	Univocal tag for reception date of the order	Datetime object "DD-MM-YYYY"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Time reference for ALL Timeseries of data (Index)
Stock Keeping Unit (SKU)	MH	Univocal tag for single product Group products by Category based on orders frequency	String "ABC012#"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Filter Signal to perform forecasts by product; and/or group of products
Client ID	MH	Univocal code/name of client for each order	Int OR String "ABC012#" --> Bin "0;1;2"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Additional KEY to merge signal and exogenous data in pre-processing phase. IF KEPT: Custom grouping and encoding ELSE: Dropped
Ordered Quantity	MH	Quantity of products ordered in uniform metric	Int/Float	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	SIGNAL forecasted



Ordered Units	O	Lots of products ordered	Int	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	IF Ordered Quantity Normalized: alternative Signal forecasted ELSE: Dropped or substitute Ordered Quantity
SALES Exogenous Features						
Order Date	MH, if SALES exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
SKU	MH, if SALES exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
Client ID	MH, if SALES exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
Expected Delivery Date	C	Univocal tag for date of delivery of ordered product	Datetime object "DD-MM-YYYY"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	To be compared with Sale Date or Order Date
Sale Date	C	Univocal tag for date of ACTUAL SALE of ordered product	Datetime object "DD-MM-YYYY"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	WARNING: Can refer only to PART of the ordered quantity
Actual Sold Quantity (AND/OR Units)	I	Quantity of product sold for each order	Float	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	WARNING: Can correspond to PART ordered quantity (Order Date VS Sale Date)
Sales Channel	I	Univocal code for sales channel of each product	String "Channel1" --> Bin "0;1;2"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Potentially higher importance if retail products To be processed with one-hot encoding
Sales Agent	C	Univocal code/name of salesperson for each product/client	String "Miguel01" --> Bin "0;1;2"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	To be processed with one-hot encoding
Unitary Price	BD	Price of single unit of product for each order	Float	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Can be relevant if it varies a lot over time

Payment timing	O	Number of days till actual payment for each order	Int	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Can be relevant if it varies a lot over time
STOCK Exogenous Features (only if MAKE TO STOCK business)						
Order Date	MH, if STOCK exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
SKU	MH, if STOCK exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
Client ID	MH, if STOCK exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
Stockout Quantity	BD	Actual demanded quantity not satisfied, net of Safety Stock	Int/Float	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Quantity of ordered SKU in Stockout on Order Date. By construction can include lot of NaNs values.
MARKETING Exogenous Features						
Order Date	MH, if MARKETING exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge+-
SKU	MH, if MARKETING exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
Market Brand	O	Brand name of for each product	String "ABC012#" --> Bin "0;1;2"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Brand power of specific product ordered can influence its demand
Product Family	O	Univocal set of products distinguished by business specific criteria	String "ABC012#" --> Bin "0;1;2"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Different products and brands can be grouped as a Family for different clients
Promotion Date Start	C	Univocal Tag for START date of planned marketing efforts relative to product SKU	Datetime object "DD-MM-YYYY"	INTERNAL database (SQL) OR records (CSV, Excel files)	IF NOT ESTIMATED : Database connector AND/OR Files	To be compared with Order Date and Promotion Date End

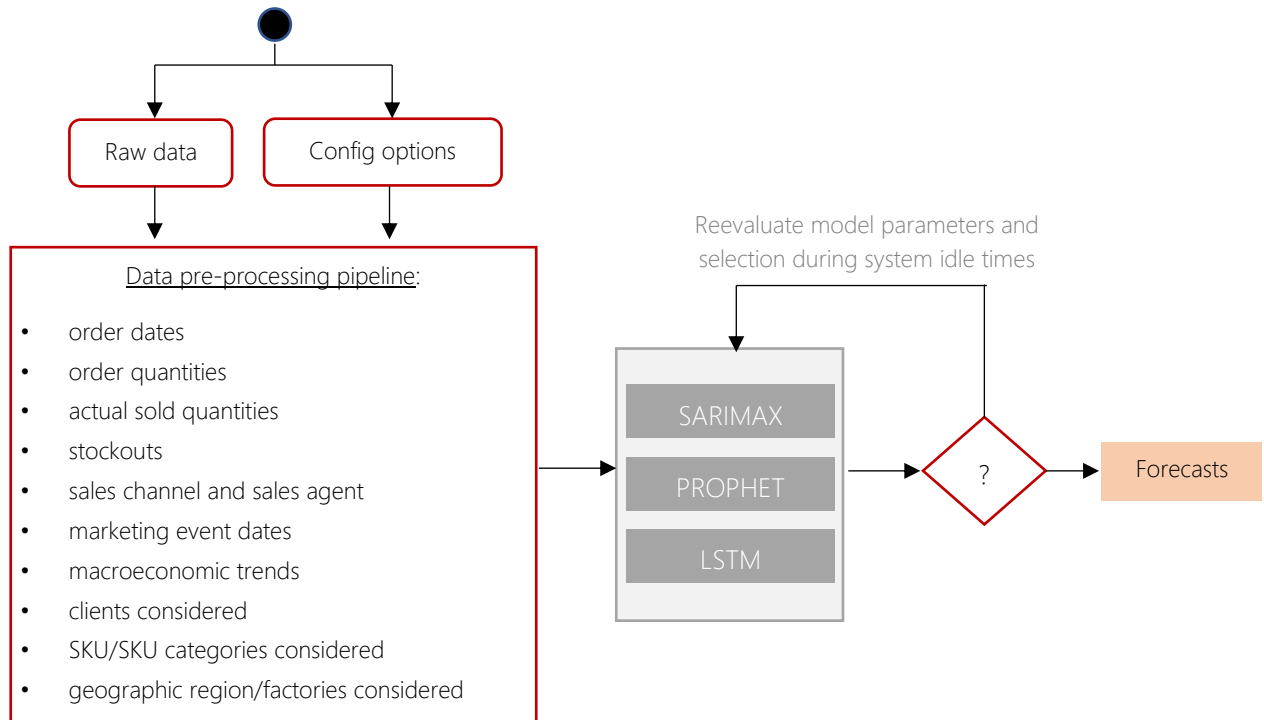
				OR MANUAL (user) OR Estimated from data	processing OR User interface calendar/d ashboard	
Promotion Date End	C	Univocal Tag for END date of planned marketing efforts relative to product SKU	Datetime object "DD-MM- YYYY"	INTERNAL database (SQL) OR records (CSV, Excel files) OR MANUAL (user) OR Estimated from data	IF NOT ESTIMATED : Database connector AND/OR Files processing OR User interface calendar/d ashboard	To be compared with Order Date and Promotion Date Start
Promotion Type	C	Type of promotion for product SKU(price discount; free shipping; premium version, etc.)	String "PromType "	INTERNAL database (SQL) OR records (CSV, Excel files) OR MANUAL (user)	Database connector AND/OR Files processing OR User interface calendar/d ashboard	To be one-hot-encoded
Promotion Amount/Q uantity	I	Discount rate on price; quantity of premium version; quantity; quantity of product SKU	Int OR Float	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Requires both Promotion Dates and Promotion Type in database or records. By construction it can assume a value only for quantifiable Promotion Types
Promotion Target	I	IF NOT General: Target clients for the promotions, identified with Client ID	Int OR String "ABC012#" --> Bin "0;1;2"	INTERNAL database (SQL) OR records (CSV, Excel files)	Database connector AND/OR Files processing	Requires both Promotion Dates and Promotion Type in database or records. Encoded Numerically (NOT one-hot)
Social Media Posts Number of Views	BD	Dates, type and number of views/comments for social media posts per product family/brand	Int AND String "ABC012#" --> Bin "0;1;2"	EXTERNAL (web) source	Custom API/web- scrapping	Can be a good alternative to Promotion, but NOT FOR PROTOTYPE
Company/ Brand Web Visibility Metrics	O	Number of visitors, clicks, search appearances etc. Per order date	Int	EXTERNAL (web) source	Custom API/web scrapping	Depending on type of business can be complementary source of info
MACROECONOMICS Exogenous Features (% variation can apply for each feature)						



Order Date	MH, if MACROECO NOMICS exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
SKU	MH, if MACROECO NOMICS exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
Client ID	MH, if MACROECO NOMICS exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
Client Location	I	GPS coordinates OR post-code OR province/region/st ate/country of client for each order	GeoSeries “(Lat Long)” OR String “Country” --> Bin “0;1;2”	INTERNAL database (SQL) OR records (CSV, Excel files) OR EXTERNAL(web) source	Database connector AND/OR Files processing OR Custom API/web- scrapping	IF GPS coordinates: treated as GeoSeries ELSE: One-hot encoding
GDP(s)	I	Gross Domestic Product estimate for PT AND client's location on order date	Float	INTERNAL source or EXTERNAL (web) source	Files processing OR Custom API/web- scrapping	Good proxy to estimate general health status of economy in a country
Industrial Index(es) Value	I	Real production output of manufacturing, mining and utility sector for PT AND client's location on order date	Float	EXTERNAL (web) source	Custom API/web- scrapping	Good proxy of health status of manufacturing sector as a whole in a country
Balance of Trade Value(s)	BD	Difference between Exports and Imports for PT AND client's location on order date	Float	EXTERNAL (web) source	Custom API/web scrapping	IF NEEDED: Good proxy for direction and amount of overall trade relationships between 2 countries
Stock Market Index(es) Value	I	Value of specific(s) reference market index(es) AND client's location on order date	Float	EXTERNAL (web) source	Custom API/web- scrapping	IF available AND relevant can provide proxy of general market trend ELSE: Major listed competitors could also be used



Commodity price	BD	Price of some raw materials of ordered product SKU on order date IF listed on commodity exchanges	Float	EXTERNAL (web) source	Custom API/web scrapping	IF available AND relevant can provide proxy of some production costs
Exchange Rate(s)	BD	Exchange rate between currencies of countries of production and of sale on each order date	Float	EXTERNAL (web) source	Custom API/web scrapping	IF available AND relevant can help forecast orders from other countries
Unemployment Rate(s)	I	Value of unemployment rate for PT AND client's location on order date	Float	EXTERNAL (web) source	Custom API/web scrapping	Good proxy to estimate domestic purchasing power of country
OTHER Exogenous Features						
Order Date	MH, if MARKETING exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
SKU	MH, if MARKETING exogenous variables are used	See Top	See Top	See Top	See Top	KEY for merge
Weather Data	BD	Values for different weather parameters for each order date (mm rain; prob rain; irradiance; wind speed; air pressure etc.)	Float	EXTERNAL (web) source	Custom API/web scrapping	Relevant for seasonal/weather dependant businesses



Model Requirements

- Three prediction models will be considered and built into the forecasting module:
 - SARIMAX;
 - Prophet;
 - LSTM.
- Of the three, SARIMAX is considered the default.
- During normal program execution, only one model will be applied to the input data. During system implementation, or should the client select the corresponding option in the initial setup screen, more than one model can be executed simultaneously. At the end of this multi-model run, a decision dashboard will be displayed, with performance metrics for the multiple models, allowing the selection of the model to be used in standard executions.
- In the initial setup screen, will also be an option for the client to define an error target, which will terminate the execution if reached.
- The models will be hypertuned automatically, via GridSearch, given a set of expected ranges that their parameters may take. The client can also adjust them manually in the decision dashboard.
- During system idle times, the forecasting module may run in the background to optimize the parameters of the models and re-evaluate the standard model selection.

- The forecasts produced by the model will be accessible through a custom API that allows the data to be exported in a set of predefined formats. Online visualizations will also be made available.

Table 2 – Summary of considered models.

Model	Comments
SARIMAX	The SARIMAX model was preferred over simpler AR/MA/ARIMA variants because exogenous variables can play a significant role in accounting for extraordinary variations.
Prophet	Prophet works best with time series that have strong seasonal effects and several seasons of historical data. It is robust to missing data and shifts in the trend, and typically handles outliers well.
LSTM	LSTM may be problematic because of the amount of data required to train a good model. Its use will be very dependent on the details of specific cases.

Performance Evaluation Requirements

- The initial configuration screen of the module will allow the client to choose the model performance evaluation metrics they want to see, from among a predefined set, in the final decision dashboard, to select the model that produces the best results for their business requirements.
- Models will be evaluated by comparison with the accuracy of the test data. The following metrics will be applied to score results:

- MAPE (Mean Absolute Percentage Error):

$$MAPE = \frac{1}{n} \sum \left(\frac{|actual - forecast|}{|actual|} \right) \times 100$$

- WAPE (Weighted Absolute Percentage Error):

$$WAPE = \frac{\frac{|actual - forecast|}{|actual|}}{\sum actual} \times actual \times 100$$

- The decision dashboard to select the final model will present the model scores and plots of the forecasted values over time, comparing them to past historic values.

User Interface

The User Interface will be composed of three main screens:

1. Initial screen, with two major sections:
 - a. Input definition:
 - i. Source of inputs (editable);
 - ii. Selection of exogenous variables to consider in the forecast (editable);
 - iii. Exogenous variables to include as estimates (editable);
 - iv. Time frame of the forecast (editable);
 - v. Option to run multiple models in parallel (editable, enables section b);
 - b. Model configuration:
 - i. Model(s) to run (editable);
 - ii. Defined error target (editable);
 - iii. Predefined evaluation metrics to display (editable);
2. Model decision dashboard (optional):
 - a. Per model, display:
 - i. MAPE;
 - ii. WMAPE;
 - b. Recommended model;
 - c. Option to rerun forecast with adjusted inputs (leads to initial screen);
 - d. Option to save model selection to be used in future runs;
 - e. Button to move forward to Final Results;
3. Final results screen:
 - a. Visualization of forecast over historic data;
 - b. Result export options;
 - c. Option to rerun forecast with adjusted inputs (leads to initial screen).