# Sustainable Development Goal Attainment Prediction: A Hierarchical Framework Using Time Series Modelling*

**Yassir Alharbi, Daniel Arribas-Be and Frans Coenen**
The University of Liverpool , Liverpool, L69 3BX, UK
{yassir.alharbi,darribas,coenen}@liverpool.ac.uk

## Abstract

A framework is presented which can be used to forecast weather an individual geographic area will meet its UN Sustainable Development Goals, or not, at some time $t$. The framework comprises a bottom up hierarchical classification system where the leaf nodes hold forecast models and the intermediate nodes and root node "logical and" operators. Features of the framework include the automated generation of the associated taxonomy, the threshold values with which leaf node prediction values will be compared and the individual forecast models. The evaluation demonstrates that the proposed framework can be successfully employed to predict whether individual geographic areas will meet their SDGs.

## 1 Introduction

In the year 2000, after a decade of conferences and summits, leaders of the world reached a consensus, to adopt a set of eight Millennium Development Goals (MDGs) [United Nations Development programme, 2007]. The eight goals were directed at different aspects of humanitarian well being. Five years later, In 2015, the success of the MDGs initiative propelled the United Nations (UN) to propose a further set of seventeen goals that they termed Sustainable Development Goals (SDGs), with an attainment date of 2030. A series of targets and indicators were identified and listed in the United Nations (UN) "Transforming our World: the 2030 Agenda for Sustainable Development" [UN, 2015]. An individual SDG, is met if the associated indicator values meet some condition. This paper presents a framework for predicting whether a given geographical region such as a continent, country or island will meet its SDGs by a given date $t$ with reference to the UN SDG dataset, a publicly available data set which at time of writing (2019) comprised $1,083,975$ records.

Whether a country meets its SDGs or not is dependant on whether individual SDGs are met, which in turn depends on whether the component targets making up an individual SDG are met, which depends on whether particular indicators, sub-indicators and, in some cases, sub-sub-indicators are met. Unlike established hierarchical classification systems, which work in a top down manner [Silla and Freitas, 2011], the envisaged prediction mechanism would work in a bottom-up manner. In both cases, the objective is to establish the class of an entity with respect to some predefined hierarchical taxonomy, and in both cases, the classification operates in a level-by-level manner. However, the branches in the top down taxonomy represent dis-junctions, whereas the branches in the bottom up case represent conjunctions. In the top down case, the identified path in the hierarchy from the root node to the leaf node holds the labels to be assigned to the entity to be classified; this is illustrated in Figure 1(a) where a classification path is highlighted. In the bottom-up case, labels associated with the leaf nodes need to be established before labels associated with parent nodes can be established, all the way up to the root node (Figure 1(b)). The taxonomy in the case of bottom up hierarchical classification can thus be thought of as a "dependency tree" [Zhang *et al.*, 2018]. An alternative way of differentiating the two approaches is to describe top down hierarchical classification as adopting a "coarse-to-fine" classification approach, whilst bottom up hierarchical classification adopts a "fine-to-coarse" classification approach. It should also be noted that top-down hierarchical classification was originally proposed as a mechanism for addressing classification problems that featured a large number of classes. It's noteworthy to mention that top down hierarchical classification techniques are well established, whereas bottom up hierarchical classification techniques are relatively under-studied.

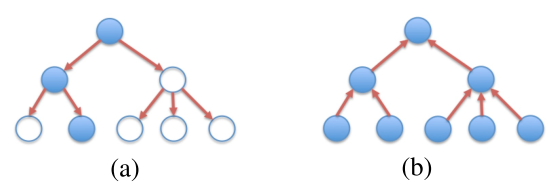



Figure 1: Hierarchical Classification; (a) Top Down, (b) Bottom Up.

In the proposed bottom up framework, each node will hold a time series forecasting model. At the root and intermedi-

*These match the formatting instructions of IJCAI-07. The support of IJCAI, Inc. is acknowledged.

ate nodes, the models will simply take binary input from their child nodes and apply a Boolean function to this input, passing the result to their parent node (or as output in the case of the root node). At the leaf nodes, the classification models will be more sophisticated addressing individual indicators, sub-indicators or sub-sub-indicators. The question to be addressed is then the nature of the forecasting models to be held at the leaf nodes. At their simplest, such models would consider a single indicator (sub-indicator or sub-sub-indicator), operating on the assumption that there is no link between the indicator and other indicators.

The rest of this paper is organised as follows. In the following section, Section 2, a brief literature review of the previous work underpinning the work in this paper is presented. The SDG data set is described in further detail in section 3. The proposed SDG bottom-up hierarchical classification framework is presented in section 4. The evaluation of the proposed framework is discussed in section 5. The paper concludes with a summary of the main finding and proposed directions for future work in section 6.

## 2 Literature Review

In this section, a brief literature review of the work underpinning the SDG prediction framework proposed in this paper is presented. The literature review commences, Sub-section 2.1 with a review of existing work directed at the SDG challenge. The problem is essentially a time series forecasting problem; hence a review of time series forecasting is presented in Sub-section 2.2. As noted in the introduction to this report, the SDG problem can be couched in terms of a Hierarchical classification problem. Hierarchical classification is therefore di

scussed in some further detail Sub-section 2.3.

### 2.1 Sustainable Development Goal Challenge

Many studies on forecasting SDGs and related challenges have been published. To monitor the progress of SDGs, the UN publishes a yearly report [UN, 2018] to measure the progress towards the global attainment of the SDGs and provides a good annual general overview. The UN also publishes statistics used to monitor progress towards SDG attainment[1]; this is the input data used with respect to the proposed framework and is therefore discussed in further detail in Section 3. The majority of the available literature has focused on forecasting individual SDGs as opposed to the whole set of SDGs as proposed by the UN [UN, 2015]. Cuaresma et al. [Crespo Cuaresma *et al.*, 2018] considered the SDG "End poverty in all its forms everywhere" (SDG Goal 1), their proposed forecasting mechanism was based on a single criteria GDP (Gross Domestic Profit) and it utilised regression-based estimates. In Shumilo et al [Shumilo *et al.*, 2018], the SDG "Life on land" (SDG Goal 15) was considered. Here the proposed forecasting mechanism was founded on the utilisation of satellite imagery by implementing neural networks to classify forest area. SDG Goal 11 was considered in [Anderson *et al.*, 2017] using data obtained from air quality sensors installed on data collection satelites.

[1] https://unstats.un.org/SDGs/indicators/database/

### 2.2 Time series forecasting

Time series analysis has been the subject of much research [Konar and Bhattacharya, nd; Hyndman, 2018]. Much of this work has been directed at supervised learning, the mapping of time series to class labels of some kind [Bagnall *et al.*, 2016]. Many methods have been proposed to predict (forecast) future occurrences in a time series data, examples include: Vector Autoregression [Stock and Watson, 2001], Holt Winters Exponential Smoothing [Gelper *et al.*, 2010] and autoregressive [Gooijer and Hyndman, 2006]. In the context of SDG prediction, a particular challenge is the nature of the time series data available; at time of writing (2019) this was limited to 18 observation points per time series.

Any forecasting method, considered in the context of the proposed framework, must therefore be able to operate using such short time series. From the literature there are three models that seem appropriate: (i) Auto-Regressive Moving Average (Arma) [Lawrance and Lewis, 1980], Auto-Regressive Integrated Moving Average (ARIMA) [Hyndman, 2018], and Facebook Prophet (Fbprophet) [Sean J and Benjamin, 2017]. Each is discussed in further detail below.

The ARMA model combines autoregressive [Mills, 1990] with a moving average model. It can be expressed as shown in equation 1, where $\phi =$ is the auto regressive models parameter, $\theta =$ is the moving average, $c =$ is a constant and $\varepsilon =$ is the error terms.

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{1}$$

The ARIMA time series forecasting model is a generalisation of the the ARMA model [Hyndman, 2018]. It can be expressed as shown in equation 2, where $t$ is a temporal index, $u$ is the mean term, $B$ is the backshift operator, $\phi(B)$ is the autoregressive operator, $\theta(B)$ is the moving average operator, and $a_t$ is the independent disturbance or the random error.

$$(1-B)^d Y_t = \mu + \frac{\theta(B)}{\phi(B)} a_t \tag{2}$$

Fbprophet is an additive regression model, directed at nonlinear time series forecasting, developed by Facebook [Sean J and Benjamin, 2017]. Fbprophet operates by decomposing a given time series into three different components referred to as "trend", "seasonality", and "holidays" and includes an error term as shown in equation 3 where $g(t)$ is trend, $s(t)$ is the the periodic change, $h(t)$ is the seasonality effect and $\varepsilon$ is the parametric assumption. The result is a model that is robust to short time series and randomness in the observation points.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{3}$$

An alternative to the above is to consider forecasting methods directed at hierarchical time series such as those proposed in Shanika [Shanika L. *et al.*, 2018] and [Hyndman, 2018], applicable where the time series under consideration was hierarchically divided. The example given in [Athanasopoulos *et al.*, 2009] is forecasting tourism in Australia. However, given that the available SDG time series are already very short the potential for a hierarchical division of these time series is very limited and unlikely to prove successful.

One of the main disadvantages of short time series forecast model generation is that there is very little opportunity for taking the presence of noise into consideration. It is argued that, inaccuracy in time series forecasting is directly related to the amount of noise in the data [Rob J. *et al.*, 2007]. The proportion of noise in short time series is often higher than in long time series. In the context of the SDG application, it is unclear how much noise there is, or how this might be defined; it can be argued that there is no spurious data and hence no noise. Whatever the case, given a collection of short time series the interaction between the time series may be utilised, although this is not considered in this paper.

### 2.3 Hierarchical Classification

As noted in the introduction to this paper, hierarchical classification is a type of supervised learning where the output of the classification is derived from a hierarchical class taxonomy [Silla and Freitas, 2011]. There are many methods directed at top-down classification [Dangerfield and Morris, 1992] and [Edwards and Orcutt, 1969] compared to bottom-up hierarchical classification founded on a taxonomy. In [Rostami-Tabar *et al.*, 2013] a new approach, called grouped time series, is discussed. This approach is applicable given an application where the required time series forecasting is to be conducted using multiple levels of granularity. For example in a warehouse stock forecasting application where we have thousands of products arranged according to a hierarchical categorisation; not quite the same as the SDG challenge but of interest because of its hierarchical nature.

## 3 The Sustainable Development Goals Data Set

As noted above, the UN identified 17 SDGs. Each SDG has between 3 and 13 targets, and each target, in turn, has a number of indicators associated with it. In most cases, the indicators have sub-indicators, and even sub-sub-indicator [Sapkota, 2019]. A summary of the SDG hi erarchy is given in Figure 2. With reference to the figure, the time series forecast models will be held at the leaf nodes, while the remaining intermediate nodes and the root node will hold Boolean functions. The nature of these Boolean functions will depend on the nature of the node. For ease of understanding, a numbering system has been adopted to identify individual indicators, $\langle g, t, i, s1, s2 \rangle$ (goal, target, indicator, sub-indicator, sub-sub-indicator), for example the identifier $[1.1.1.1.1]$ indicates: Goal1, Target 1, Indicator 1, Sub-indicator1, Sub-sub-indicator 1.

The SDG data set is publicly available from the SDG website[2]. At time of writing (2019) the data set spanned an 18 year period. The SDG data set is relatively large, 500MB, and is comprised of some 1,100,000 records holding statistical SDG information covering individual geographic areas. The majority of geographic areas considered are countries,

[2]https://unstats.un.org/SDGs/indicators/database/

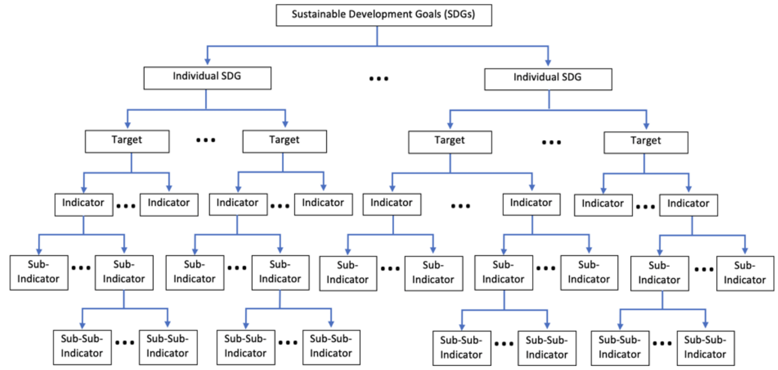



Figure 2: SDG Hierarchy

continents and islands that currently exist, 195 of them. The remainder comprise countries that currently are no longer in existence and geographic groupings of countries. Each record references a particular time stamp (year), geographical area and indicator (sub -indicator or sub-sub-indicator). The data is organised according to 36 columns (attributes) . The first three columns list the goal, target and indicator referenced by each record. The geographical area ID and name are given in columns 6 and 7 and the associated time stamp in column 8. The remaining 29 columns give additional information concerning whether a record referrers to a sub-indicator or a sub-sub-indicator or not, and relevant values. In many cases the attribute referenced by the column is not applicable, hence no value is given. For example the last attribute, column 37, refers to internet speed which is irrelevant with respect to most indicators. In other cases the the column is applicable, but the value is missing. Hence the data set features both "absent" and "missing" values"; a summary of the number of absent and missing values featured in the data set is given in Figure 3.

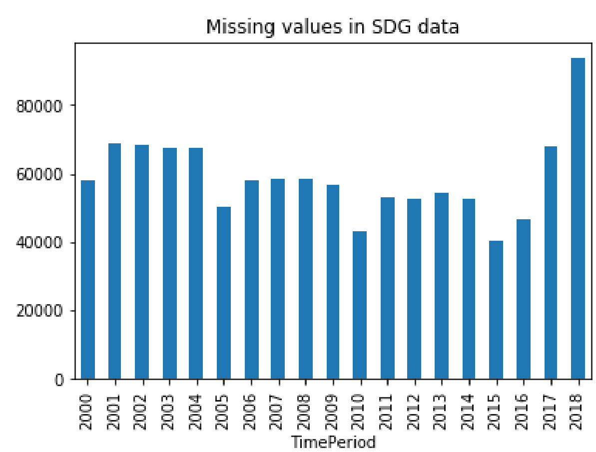



Figure 3: Histogram summarising number of SDG absent and missing data values per sample year.

As noted above the data set spans an 18 year period, thus for a given geographic area and a given indicator (sub-indicator or sub-sub-indicator) there will be a time series comprised of a maximum of 18 points (values). There are records where the time series only feature a small number of points, the remaining values being missing.

The SDG data set $D$, comprises of a single table measuring $r \times |A|$, where $r$ is the number of records and $|A|$ is the size of the attribute set (the number of columns). At time of writing $r = 1,083,975$ and $|A| = 37$. To generate the desired forecast models the data set $D$ had to be "reshaped" [Wang *et al.*, 2019] to give a data set $D' = e \times y$ where $e$ is the number of leaf nodes that will feature in the SDG hierarchy, multiplied by the number of geographic areas covered by the SDG data set, and $y$ is the number of years for which data is available. At time of writing $D' = 1803096$ $(18 \times 100172)$ and $y = 18$; it is anticipated that $y$ will increase year-by-year as further data becomes available. The data set $D'$ holds numeric values only. In effect each row in $D'$ is a time series $\{v_1, v_2, \ldots, v_y\}$ which in turn can be used to build the desire forecast models.

# 4 The SDG Prediction Framework

There are two aspects to the Prediction Framework: (i) the generation of the taxonomy and associated constraints to be embedded in the framework, a generic process independent of the geographic region of interest; and (ii) prediction model generation, a geographic region dependent process that will be repeated for each geographic region to be considered. Each is discussed in further detail in the following three subsections. A schematic of the proposed process is given in Figure 4. In the proposed method the raw data, coming from the UN repository, is processed to produce a modelling ready data set and a hierarchical taxonomy with constraints.

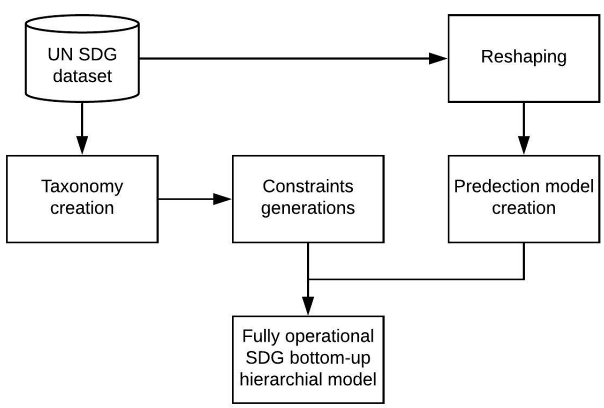



Figure 4: System overview.

## 4.1 SDG Taxonomy Generation

Hierarchical classification (top-down or bottom-up) requires a taxonomy and associated hierarchy. In many cases of top-down hierarchical classification, the hierarchy and taxonomy are easily defined and are often quite trivial. In the case of the SDG hierarchy, the hierarchy and taxonomy are substantial as indicated in Figure 2. Further, the UN does not provide a taxonomy for the data. Therefore the taxonomy and hierarchy need to be extracted from $D$ (the UN SDG data set). Hand-crafting of the taxonomy and hierarchy was clearly not a desirable option, as it would be time-consuming and prone to error; there is also the potential that the UN may change elements of the SDGs, or add a completely new goal or edit an existed one. An automated approach to generating the taxonomy and hierarchy and was therefore seen as desirable. A Hierarchical Taxonomy Generator was developed for this purpose, the input for which was the raw SDG data for all geographical regions. This was developed using the Python Pandas library for data manipulation and analysis, specifically the cross-tabulation (*Crosstab*) function defined in the Pandas library. This allowed for the automated generation of a SDG taxonomy from $D$ from which the associated hierarchy could be inferred.

## 4.2 Threshold Generation

Each node in the hierarchy has a boolean condition associated with it. At the the root and intermediate nodes the condition are expressed simply as a "logical and"; if all the inputs have the value $True$ the output value is $True$, and $False$ otherwise. At the leaf modes the conditions are more complex and are outlined in the SDG Handbook [Sapkota, 2019]. These are typically expressed in the form of some conditional operator, such as greater than ($>$), less than ($<$) or equal to ($=$), some predefined threshold $\sigma$. The challenge is that the $\sigma$ values to be associated with the leaf nodes are not included in $D$ and are not specified in [Sapkota, 2019]. Instead, they are published separately in [UN, 2017]. However, in [UN, 2017] some of the thresholds are not mathematically defined. A solution, in the context of the proposed hierarchical framework, was available in the [Lozano *et al.*, 2018] where the authors published guidelines on how to interpret the health target goals from the SDG published Target goals document, including mathematical definitions. The generated thresholds could be related directly to the SDG Taxonomy produced by the Hierarchical Taxonomy Generator. Once the SDG Taxonomy has been generated it was ready for use to automatically generate the SDG prediction hierarchy.

## 4.3 Forecast Model Generation

As noted above, each leaf nodes in the hierarchy will hold a forecast model. The forecast models at the leaf nodes are required to predict what the value associated with the indicator in question will be and then to determine whether that value meets its specified threshold value $\sigma$ or not. However, unlike the prediction hierarchy, generated as described above, the nature of the forecast models are specific to individual geographic regions and thus each needs to be generated on a "as required" basis. The forecast models held at the leaf nodes were generated using the available data for each indicator (sub-indicator or -sub-sub-indicator) associated with each geographic area included in the SDG data set, over 100,000 of them. A number of forecast model generation mechanism were considered: (i) Auto Regression Moving Average (ARMA) [Lawrance and Lewis, 1980], (ii) Auto-Regressive Integrated Moving Average (ARIMA) [Kinney, 1978] and

Table 1: Framework evaluation using Target 3.2 and the geographic area Egypt

| Goal | Target | Arima $RMSE$ $MAPE$ | Arma $RMSE$ $MAPE$ | Fbprophet $RMSE$ $MAPE$ | Series Description | Series Code | Initial Value | Prediction | Threshold value | Date | Result |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 3.2 | 0.591 4.410% | 5.349 41.260 | 0.016 0.079% | Neonatal mortality rate | bothsex 1m | 12.5 | 13.172 | =12 | 2030 | Not met |
| | | 8432 6.197% | 19975 16.130% | 2755 1.852% | Under-five deaths | male 5y | 32537 | 35278 | $\leq$25% | 2030 | Not met |
| | | | | | | bothsex 5y | 59728 | 63777 | $\leq$25% | 2030 | Not met |
| | | | | | | female 5y | 27191 | 30430 | $\leq$25% | 2030 | Not met |
| | | 5115 4.475% | 14258 13.376% | 2688 2.188% | Infant deaths | male 1y | 27957 | 31526 | $\leq$25% | 2030 | No met |
| | | | | | | bothsex 1y | 50924 | 57755 | $\leq$25% | 2030 | Not met |
| | | | | | | female 1y | 22967 | 24871 | $\leq$25% | 2030 | Not met |
| | | 2190 6.339% | 5423 16.472% | 66.095 0.153% | Neonatal deaths | bothsex 1m | 31796 | 32688 | $\leq$25% | 2030 | No met |
| | | 1.015 1.219% | 31.661 43.846% | 0.010 0.006% | Under-five mortality rate , by sex | male 5y | 25.1 | 25 | $\leq$25% | 2030 | Not met |
| | | | | | | bothsex 5y | 23.7 | 25 | $\leq$25% | 2030 | Not met |
| | | | | | | female 5y | 22.3 | 26 | $\leq$25% | 2030 | Not met |
| | | 0.771 1.121% | 24.000 0.392% | 0.016 0.012% | Infant mortality rate | male1y | 21.4 | 23 | $\leq$25% | 2030 | Not met |
| | | | | | | bothsex 1y | 20.1 | 21 | $\leq$25% | 2030 | Not met |
| | | | | | | female 1y | 18.7 | 20 | $\leq$25% | 2030 | Not met |

(iii) Facebook Prophet (Fbprohphet) [Sean J and Benjamin, 2017].

# 5 Evaluation

The evaluation of the proposed framework is presented in this section. The evaluation comprised two elements: (i) evaluation of the forecast models generation mechanism, (ii) evaluation of the the framework as a whole.

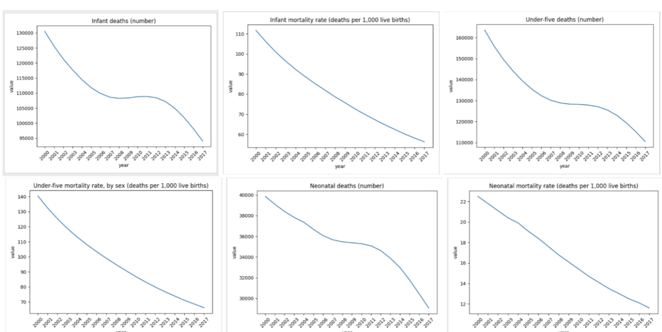
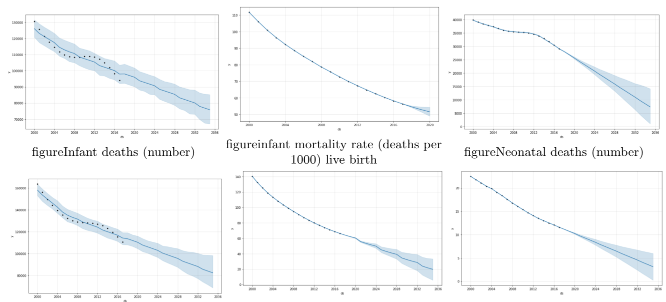



Figure 5: Raw time series associated with the six indicators for SDG Target 3.2.

## 5.1 Forecasting evaluation

As noted above, three forecast model generators were considered: (i) ARMA, (ii) ARIMA and (iii) Fbprohphet. The evaluation metrics used were: Root Means Square Error (RMSE) and Means Absolute Percentage Error (MAPE) [Hyndman and Koehler, 2006]. RMSE is calculated as shown in equation 4 where $f$ is the forecasted value and $o$ is the observed value. RMSE provides results with the same unit as the forecasted values and therefore it is easy to compare the result of two forecasting methods; however, the metric is not an intuitive one. MAPE is calculated as shown equation 5 where $f$ is the forecasted value and $o$ is the observed . MAPE offers an easy to understand forecasting percentage error, although the units are lost.



figureInfant deaths (number) figureinfant mortality rate (deaths per 1000) live birth figureNeonatal deaths (number) figureUnder-five deaths (number) figureUnder-five mortality rate, by sex (deaths per 1,000 live births) figureNeonatal mortality rate (deaths per 1,000 live births)

Figure 6: Output using Fbprohphet, indicating prediction spread, with respcet to Time series associated with the six indicators for SDG Target 3.2 prediction

$$RMSE = \sqrt{(f-o)^2} \tag{4}$$

$$MAPE(\frac{1}{n}\sum_{i=1} i = n\frac{o_i - f_i}{o_i} * 100) \tag{5}$$

For the evaluation SDG Target 3.2, “By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births”, was selected, together with the geographic area Egypt. This was selected because a complete set of data points was available. Target 3.2 comprised six indicators; the associated time series are given in Figure 5. The forecast models were trained using the first seventeen data points and used to predict the eighteenth (2018) value. The accuracy of the prediction was measured using RMSE and MAPE. The results are given in Table 1, columns 3, 4 and 5. From the table, it can be seen that the Fbprophet prediction model produced the best results. For example in the case of forecasting ”Neonatal mortality rate

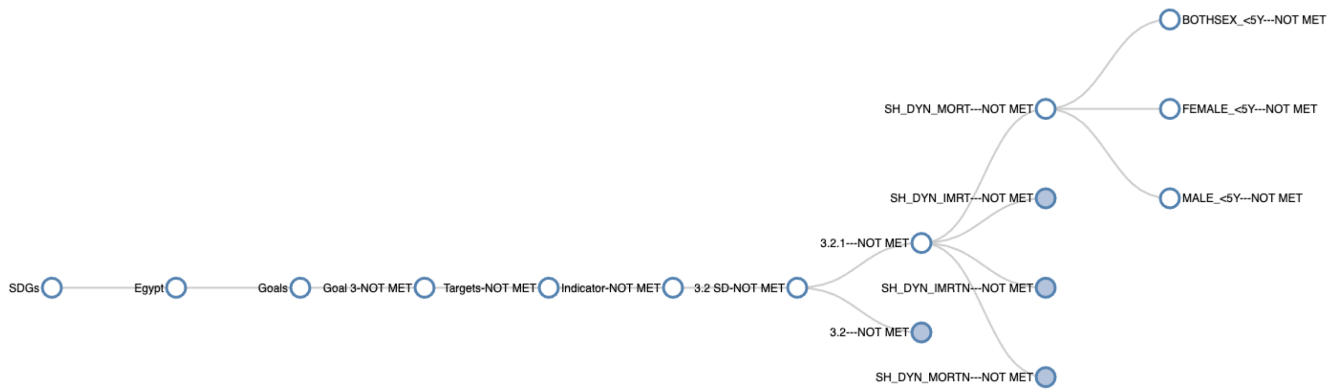



Figure 7: Example prediction visualisation for Target 3.2 with respect to the geographic area Egypt.

(deaths per 1,000 live birth") the RMSE score was $0.55$ using ARIMA, $5.24$ using ARMA and $0.016$ using Fbprophet. Figure 6 shows the output using Fbprophet.

### 5.2 Framework Evaluation

To evaluate the utility of the proposed SDG framework the geographic area Egypt was again used together with SDG Target 3.2. The framework was then used to automatically predict whether the target will be met by 2030, as specified in the UN Agenda for Sustainable Development. Target 3.2, as noted above, encompasses six indicators, six forecast models were therefore generated using Fbprophet (because earlier evaluation, reported on in Sub-section 5.1, had shown this produced best results). The prediction models were trained using the first seventeen data points and then used to predict the 2030 values which were then used to automatically determine, using the framework. whether the indicators were met, or not, by comparing the forecasted values with the appropriate threshold value. To predict whether Target 3.2 will be met in 2030 all forecasted values must be less the $25\%$ of the benchmark value for the year 2015. The results are presented in Table 1, columns 9, 10 and 12. From the table, it can be seen that in the case of the geographic area Egypt and Target 3.2 the target will not be met by 2030. However, if the "trend" for each indicator, using the first seventeen points, is examined, as shown in Figure 6, it can be seen that the forecasted value will meet the threshold value at some time in the future.

### 5.3 Framework Visualisation

An additional feature of the framework is that it includes a visualisation of predictions in the form of D3.js dendrograms [Bostock *et al.*, 2011]. [3] An example prediction visualisation for Target 3.2 with respect the the geographic area Egypt is given in Figure 7. In the Figure blue nodes are "closed nodes"; double clicking on closed nodes causes the node to "open out" to the next level.

## 6 Conclusion

A framework has been presented for predicting whether individual geographic areas will meet their UN SDGs at a given time $t$. The framework comprises a bottom up classification hierarchy where the leaf nodes hold predictors founded on time series data and the intermediate nodes and root node simple "logical and" operators. A feature of the framework is that the required hierarchical classification taxonomy is generated automatically. For individual geographic areas individual time series-based predictors are required, these are also generated in an automated manner. The framework was evaluated by considering a number of prediction models, and by using it to predict whether individual geographic areas would meet their targets by 2030 as specified in the UN Agenda for Sustainable Development. The best prediction model was found to be Fbprophet. The evaluation indicated that the proposed framework could be successfully employed to predict whether geographic areas would meet their targets or not.

[3] https://bit.ly/2K8fEcj

## References

[Anderson *et al.*, 2017] Katherine Anderson, Barbara Ryan, William Sonntag, Argyro Kavvada, and Lawrence Friedl. Earth observation in service of the 2030 agenda for sustainable development. *Geo-spatial Information Science*, 20(2):77–96, 2017.

[Athanasopoulos *et al.*, 2009] George Athanasopoulos, Roman A. Ahmed, and Rob J. Hyndman. Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting*, 25(1):146 – 166, 2009.

[Bagnall *et al.*, 2016] Anthony Bagnall, Jason Lines, , Jon Hills, and Aaron Bostrom. Time-series classification with cote: The collective of transformation-based ensembles. *2016 IEEE 32nd International Conference on Data Engineering (ICDE), Data Engineering (ICDE), 2016 IEEE 32nd International Conference on*, page 1548, 2016.

[Bostock *et al.*, 2011] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, December 2011.

[Crespo Cuaresma *et al.*, 2018] Jesus Crespo Cuaresma, Wolfgang Fengler, Homi Kharas, Karim Bekhtiar, Michael Brottrager, and Martin Hofer. Will the sustainable development goals be fulfilled? assessing present and future global poverty. *Palgrave Communications*, 4, 12 2018.

[Dangerfield and Morris, 1992] Byron J. Dangerfield and John S. Morris. Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International Journal of Forecasting*, 8(2):233 – 241, 1992.

[Edwards and Orcutt, 1969] John B. Edwards and Guy H. Orcutt. Should aggregation prior to estimation be the rule? *The Review of Economics and Statistics*, 51(4):409–420, 1969.

[Gelper *et al.*, 2010] Sarah Gelper, Roland Fried, and Christophe Croux. Robust forecasting with exponential and holtwinters smoothing. *Journal of Forecasting*, 29(3):285–300, 2010.

[Gooijer and Hyndman, 2006] Jan G. De Gooijer and Rob J. Hyndman. 25 years of time series forecasting. *International Journal of Forecasting*, 22(3):443 – 473, 2006. Twenty five years of forecasting.

[Hyndman and Koehler, 2006] Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 2006.

[Hyndman, 2018] Rob J Hyndman. Forecasting: principles and practice, May 2018.

[Kinney, 1978] William R. Kinney. Arima and regression in analytical review: An empirical test. *The Accounting Review*, 53(1):48–60, 1978.

[Konar and Bhattacharya, nd] Amit Konar and Diptendu Bhattacharya. *Time-series prediction and applications : a machine intelligence approach.* Intelligent systems reference library: volume 127. Springer, n.d.

[Lawrance and Lewis, 1980] A. J. Lawrance and P. A. W. Lewis. The exponential autoregressive-moving average earma(p,q) process. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):150–161, 1980.

[Lozano *et al.*, 2018] Rafael Lozano, Nancy Fullman, Degu Abate, Solomon Abay, Cristiana Abbafati, Nooshin Abbasi, Hedayat Abbastabar, Foad Abd-Allah, Johan Ärnlöv, and Christopher J. L Murray. Measuring progress from 1990 to 2017 and projecting attainment to 2030 of the health-related sustainable development goals for 195 countries and territories: a systematic analysis for the global burden of disease study 2017. *The Lancet*, 392(10159):091–2138, 2018.

[Mills, 1990] Terence C. Mills. *Time series techniques for economists.* Cambridge : Cambridge University Press, 1990., 1990.

[Rob J. *et al.*, 2007] Hyndman Rob J., Kostenko Andrey V., and Points Key. Minimum sample size requirements for seasonal forecasting models. *International Journal of Applied Forecasting*, 2007.

[Rostami-Tabar *et al.*, 2013] B. Rostami-Tabar, M. Z. Babai, A. A. Syntetos, and Y. Ducq. Forecasting aggregate arma(1,1) demands: Theoretical analysis of top-down versus bottom-up. In *Proceedings of 2013 International Conference on Industrial Engineering and Systems Management (IESM)*, pages 1–8, Oct 2013.

[Sapkota, 2019] Shaswat Sapkota. *E-Handbook on Sustainable Development Goals*. United Nations, 2019.

[Sean J and Benjamin, 2017] Taylor Sean J and Letham Benjamin. Forecasting at scale. *The American Statistician*, 2017.

[Shanika L. *et al.*, 2018] Wickramasuriya Shanika L., Athanasopoulos George, and Hyndman Rob J. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, 2018.

[Shumilo *et al.*, 2018] L. Shumilo, A. Kolotii, M. Lavreniuk, and B. Yailymov. Use of land cover maps as indicators for achieving sustainable development goals. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 830–833, July 2018.

[Silla and Freitas, 2011] Carlos N. Silla and Alex A. Freitas. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, 22:31–72, 2011.

[Stock and Watson, 2001] James H. Stock and Mark W. Watson. Vector autoregressions. *Journal of Economic Perspectives*, 15(4):101–115, December 2001.

[UN, 2015] United Nations UN. Transforming our world: the 2030 agenda for sustainable development. Working papers, eSocialSciences, 2015.

[UN, 2017] UN. Tier classification for global sdg indicators. *United Nation*, 2017.

[UN, 2018] UN. *The Sustainable development goals report 2018*. [New York] :. 2018, 2018. Available also online (viewed 11 July 2018).

[United Nations Development programme, 2007] United Nations Development programme. Millennium Development Goals, 2007.

[Wang *et al.*, 2019] Earo Wang, Dianne Cook, and Rob J Hyndman. A new tidy data structure to support exploration and modeling of temporal data. *arXiv e-prints*, page arXiv:1901.10257, Jan 2019.

[Zhang *et al.*, 2018] Chao Zhang, Fangbo Tao, Xiusi Chen, Jiaming Shen, Meng Jiang, Brian Sadler, Michelle Vanni, and Jiawei Han. Taxogen: Unsupervised topic taxonomy construction by adaptive term embedding and clustering. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, pages 2701–2709, New York, NY, USA, 2018. ACM.