

AN INTERACTIVE PLATFORM FOR ENVIRONMENTAL SENSORS DATA ANALYSES

Shubhi Harbola^{1,2}*, Volker Coors²

¹ University of Stuttgart, Stuttgart, Germany - Shubhi.Harbola@ieee.org

² University of Applied Sciences, Stuttgart, Germany - (Shubhi.Harbola, Volker.Coors)@hft-stuttgart.de

KEY WORDS: Transformers Network, sensor monitoring, hierarchical clustering, predictions, cities planning, visualisation, unsupervised classification, pollution parameters, machine learning, visual analytics, meteorological data

ABSTRACT:

The increased usage of the environmental monitoring system and sensors, installed on a day-to-day basis to explore information and monitor the cities' environment and pollution conditions, are in demand. Sensor networking advancement with quality and quantity of environmental data has given rise to increasing techniques and methodologies supporting spatiotemporal data interactive visualisation analyses. Moreover, Visualisation (Vis) and Visual Analytics (VA) of spatiotemporal data have become essential for research, policymakers, and industries to improve energy efficiency, environmental management, and cities' air pollution planning. A platform covering Vis and VA of spatiotemporal data collected from a city helps to portray such techniques' potential in exploring crucial environmental inside, which is still required. Therefore, this work presents Vis and VA interface for the spatiotemporal data represented in terms of location, including time, and several measured attributes like Particular Matter (PM) $PM_{2.5}$ and PM_{10} , along with humidity, and wind (speed and direction) to assess the detailed temporal patterns of these parameters in Stuttgart, Germany. The time series are analysed using the unsupervised HDBSCAN clustering on a series of (above mentioned) parameters. Furthermore, with the in-depth sensors nature understanding and trends, Machine Learning (ML) approach called Transformers Network predictor model is integrated, that takes successive time values of parameters as input with sensors' locations and predict the future dominant (highly measured) values with location in time as the output. The selected parameters variations are compared and analysed in the spatiotemporal frame to provide detailed estimations on how average conditions would change in a region over the time. This work would help to get a better insight into the urban system and enable the sustainable development of cities by improving human interaction with the spatiotemporal data. Hence, the increasing environmental problems for big industrial cities could be alarmed and reduced for the future with proposed work.

1. INTRODUCTION

The cities generate and store a lot of spatial and temporal information continuously using sensors that collect a large set of real-time spatial data stream and responses. In monitoring and keeping track of the surroundings, managing spatial data includes cities, rivers, roads, and countries with increasing demand for environmental monitoring, smart cities planning and resource management. The development and industrial advancement for uplifting human standards have contributed to a comfortable life on one hand while consequences of environmental changes, and pollution on another. Chimneys' discharge, waste from industries, vehicle smokes, and construction sites release consist of tiny air pollutants that could be inhaled with the air, leading to heart diseases, lung and respiratory problems worldwide. Therefore, the meteorological parameters *i.e.*, humidity, wind (speed and direction) along with air pollutants like Particular Matter (PM) $PM_{2.5}$ and PM_{10} require regular monitoring. The surrounding air quality and well being fluctuate with these parameters atmospheric concentrations (Chen and Zhao, 2011). Over the developed areas, the elevated levels of pollution parameters are incorporated with both local emission sources, and regional transportation (Chen and Zhao, 2011, Jasen et al., 2013). Regional transportation with diesel vehicles are the primary sources of particular matters and contribute significantly to their levels (Wallace and Hobbs, 1977, Hardin and Kahn, 1999). Moreover, sensors' (spatial and temporal) data is a combination of the georeferenced geographical entity represented in terms of location, dimensions, attributes, and time as continuous more extensive size data. Data Visualisation (Vis) is not an instrument for Visual Analytics (VA). More-

over, VA is a sub-field of Vis which integrates data analyses with highly interactive visualisations. Furthermore, in the scientific domain, it helps to place the geospatial data in a visual context by identifying trends, patterns that usually go unrecognised in the text-based data (Sun et al., 2013, Sun and liu, 2016, Harbola and Coors, 2018). Some existing studies have been performed to infer the seasonality and patterns insides for meteorological and pollution parameters independently (Garrett and Casimiro, 2011, Harbola and Coors, 2020). Integrating interactive Vis techniques help in representing the geospatial data and attached environmental information together in one frame, beyond the typical spreadsheets, charts and graphs, along with presenting it in more sophisticated formats using infographics, maps, detailed bars, pie and heat maps to communicate in between relationships (Horvitz, 2007, Aigner, 2013, Liu et al., 2017). However, VA combines automated analysis techniques with interactive Vis, thereby assisting in the easier understanding of the temporal sensors data along with decision-making capabilities by dividing the cities into several components varying over space, time, and different spatial scales (Kurkcu et al., 2017, Panagiotopoulou and Stratigea, 2017). Several above discussed studies used smoothening and filtering techniques, ignoring the data noise and modifying the originality of the temporal dataset. Interactively visualising the sensors, and their data concerning timeframe helps monitor these parameters. The comprehensive study of meteorological parameters and their contribution to $PM_{10-2.5}$, could be helpful. The above research suggests that several questions remain to be addressed, such as temporal wind variations, $PM_{10-2.5}$ concentration fluctuations and in connection with user desired time frame, without modifying the authenticity of the original temporal dataset. An interactive system, Air Quality Temporal Anal-

*Corresponding author.

user (AQTA), is developed, supporting the visual analyses of air quality data with time (Harbola et al., 2021). It discovers temporal relationships among complex air quality data, interactively in different time frames, by harnessing the user’s knowledge of factors influencing the behaviour with the aid of Machine Learning (ML) models, but on a small scale, detailed for each sensor (individually) with lacking the spatial knowledge attached. A better insight into the sensor’s system by improving human interaction with recorded measurements and spatiotemporal information is still required. This motivates the current research. The climate fluctuation and meteorological data monitoring concerns increased the demand for such a Web interface to study the measured data history interactively along with the sensors’ nature monitoring for the future. This idea is implemented and expanded as a case study for Stuttgart (Germany).

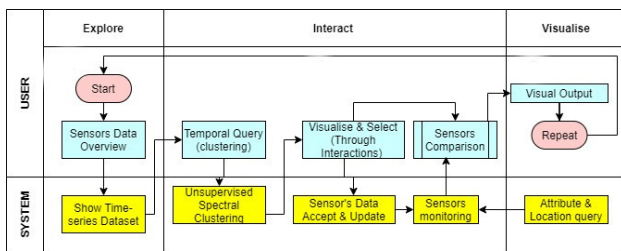


Figure 1: Sensors visual comparative analyses flowchart.

Thus, unsupervised Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) based clustering algorithm and Machine Learning (ML) model using Transformers Network designed for sensors nature monitoring and highlighting the dominating sensors locations working on the original temporal datasets by taking into consideration the above-listed gaps, and addressing solutions. These frameworks combine together to form (shown in Figure 1) Environmental Sensors Visual Prediction Assessment (ESVPA) an interactive visualisation platform with user choice parameters’ selection freedom, delivering temporal variations of spatiotemporal information. Therefore, the current study proposes HDBSCAN clustering and sensors nature monitoring queries with the following contributions: (i) interactive temporal visualisation of unsupervised cluster identifications to support the user in the interpretation of the meteorological and pollution parameters, (ii) predicting sensor nature using Transformers Network, supported with visualisation of designed model dynamic training, testing and accuracy metrics assessments. Thereby highlighting the respective model’s success and failure for inference data, (iii) visual preservation of spatial, non-spatial context and historical dataset information on user-selected temporal frame, and (iv) Unboxing the complexities of ML design with visualisation to making concept understanding more explainable and straightforward. This interactive visualisation platform would help to infer smart decisions for surrounding quality planning, which would help in proficient management and development of the city’s resources. The remaining paper is organised as follows: Section 2. presents the methodology, datasets, and results, including discussion are explained in Section 3. and Section 4., respectively, followed by the conclusion in Section 5..

2. METHODOLOGY

The proposed interactive web interface provides a platform to view and analyse in detail several sensors and their measurements in Stuttgart city along with spatial and temporal information. Each of the sensors are measuring parameters like PM_{2.5} and PM₁₀, humidity, wind (speed and direction). Following sec-

tion explains proposed system architecture comprising of unsupervised HDBSCAN clustering in 2.1, sensors nature prediction using Transformers Network in 2.2, and interactive visualisation platform inside in 2.3.

2.1 Unsupervised HDBSCAN Clustering

All sensors time series measurements (for each sensor location) are studied using unsupervised clustering and sensor’s location queries. Initially, values of each parameter are preprocessed before applying the clustering. The preprocessing involves normalising of the data followed by temporal filtering. The mean and standard deviation of a parameter are calculated. The values of a parameter are then subtracted by mean, followed by division with standard deviation, to get the normalised value. Further, the temporal filtering is applied on these normalised values. In the current study, the interactive temporal filtering based on user selection in a years is applied. These user selection temporal query division helps in detailed analysis of the considered parameters as per user desires. HDBSCAN is applied in this study on sensors’ measurements with noise which is an extension of Density-Based Spatial Clustering (DBSCAN) by converting it into a hierarchical clustering algorithm. It performs DBSCAN over varying epsilon (eps) values (*i.e.*, eps-neighborhood of point X, defining the radius of neighborhood around a point X) and integrates the results to find a clustering that gives the best stability over eps (Campello et al., 2013). This allows HDBSCAN to find clusters of varying densities (unlike DBSCAN). Therefore, for historical data HDBSCAN returns good clusters with little parameter tuning. The minimum cluster size parameter, is intuitive and decided empirically in this study. Values of k_1 and k_2 (Table 1) were empirically taken as 0.75 and 0.35 respectively (same for all the parameters), so that a sufficient number of samples occur in each class (as shown in Table 1).

Thus, HDBSCAN is one of the strongest clustering option with these advantages, it is applied on the temporal measurements of sensors’ and produce interactive filtering output. The generated distance matrix in hierarchical clustering helps in identifying the similarities of the clusters and combines most similar clusters hierarchically until the desired number of clusters are obtained minimising the variance within the cluster by using the objective function of the error sum of squares (McInnes and Healy, 2017). The sum of the squares starting from the designed three clusters (low, mild and high) is kept minimised. This gives a hint through the merging cost. The number of clusters is kept fixed until the merging cost increases and then used the cluster (value ranges), right before the merging cost increased simultaneously (Paul and Murphy, 2009).

2.2 Transformers Network

In order to provide more detailed comparison and trends analysis, each sensors’ nature monitoring using Machine Learning (ML) approach called Transformers Network predictor model were designed and integrated, that takes successive time values in terms of parameters as input with sensors’ locations and predict the future dominant (high measurements) value and location with time as the output (Wu et al., 2020). A Transformers Network is an encoder-decoder architecture, here the encoder consists of some set of encoding layers that process the input iteratively, one layer after another, and the decoder consists of a group of decoding layers that do the same thing to the encoder’s output (Vaswani et al., 2017). Each encoder layer’s function is to process its input to generate encodings, containing information about which parts of the inputs are relevant to each other. It passes its set of encodings to the next encoder layer as inputs. Each decoder

Table 1: Various classes designed ranges.

Class	Lower Range	Upper Range
1	$\min(selected_{parameter})$	$\mu - k_1\sigma$
2	$\mu - k_1\sigma$	$\mu + k_2\sigma$
3	$\mu + k_2\sigma$	$\max(selected_{parameter})$

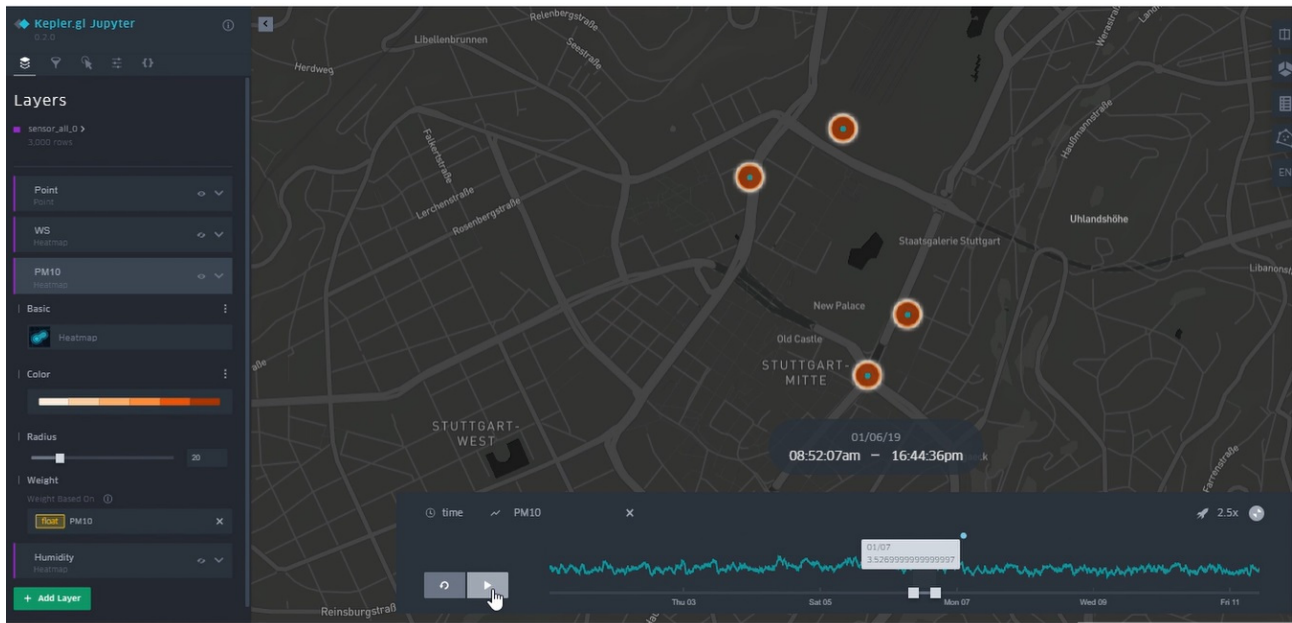


Figure 2: PM₁₀ concentrations measured by the sensors on map Web interface.

layer does the opposite, taking all the encodings and processes them, using their incorporated contextual information to generate an output sequence. Each decoder layer also has an additional attention mechanism that draws info from previous decoders' outputs before the decoder layer draws data from the encodings (Parmar et al., 2018). The encoder and decoder layers have a feed-forward neural network for additional processing of the outcomes and contain residual connections and layer normalization steps.

The used dataset comprises of wind (speed and direction), humidity, PM_{2.5} and PM₁₀, with temporal resolution $epoch$ and $epoch_j$ ($j \rightarrow 1$ to n) denotes wind (speed and direction), humidity, PM_{2.5} and PM₁₀ at time j , where 1 and n are the first and last values in the dataset, respectively. Multiple samples are designed using the dataset for training and testing the proposed algorithms. A sample consists of a feature vector as an input with a corresponding three output classes. $Window_b$ (a scalar) consecutive values of wind (speed and direction), humidity, PM_{2.5} and PM₁₀ from $epoch_j$ to $epoch_{j+Window_b}$ form a feature vector of dimension $Window_b \times 1$ which is the input of the sample for each parameter. $Window_f$ (a scalar) successive values of considered five parameters after the last value in the input *i.e.*, $epoch_{j+Window_b}$, are used to define the sample's output class. Mean (μ), and standard deviation (σ) of the wind (speed and direction), humidity, PM_{2.5} and PM₁₀ of the entire dataset are calculated. Various class boundaries are designed using μ and σ as shown in Table 1. Among $Window_f$, count of values occurring in each class in Table 1 is noted, and the class that has a maximum count *i.e.*, dominant, is assigned to the sample. Similarly, multiple samples based on each of the parameters are created by taking $Window_b$ values in the corresponding input from $epoch_j$ to $epoch_{j+Window_b}$ by varying j from 1 to $n - Window_f$, at an increment of 1. The outputs of these samples are designed as discussed above. Thus, at this stage, for $Window_b$ values in the input from $epoch_j$ to $epoch_{j+Window_b}$, there would be five sets

of samples, one based on humidity, wind speed, wind direction and other based on PM_{2.5} and PM₁₀. Here in this analysis the size of $Window_b$ and $Window_f$ are kept equal with user option to predict the next 6 hours. These conditions ensured comprehensive and accurate analysis of the data with respect to independent and different parameter selections.

2.3 Visualisation Platform

Moreover, an interactive platform is developed to provide an in depth analytic and nature patterns clarity in between the meteorological and pollution parameters for user desired inputs in the desired time frame. This platform is called as Environmental Sensors Visual Prediction Assessment (ESVPA) for sensors nature monitoring. ESVPA also provides tooltipping, brushing and linking for maintaining the transparency and combining different visualisation methods between user-computer efficient interactions (Shneiderman, 1996, Horvitz, 2010). Figure 1 provides an overview of ESVPA workflow, with highlighting the system-user interfaces of visual sensors prediction and analyses. The *System* combines with historical meteorological and pollution parameters temporal database, unsupervised clustering outputs, sensor nature monitoring Transformers Network, structure of various graphs and charts, and accepts user queries. The *User* raise queries, selects, inspects and views the states of the parameters interactively.

ESVPA uses a time series dynamic stack chart with a calendar selection to visually interact with the parameters and spatial information attached with the help of the interactive map. This is accompanied with map to provide a detailed time series data inspection of parameters magnitudes with line chart, calendar chart and heat-map view option in order to compare the trends among Sensors' parameters based on the months during a year, week, day-wise. Figure. 2, and Figure. 3 provides a glimpse of the de-

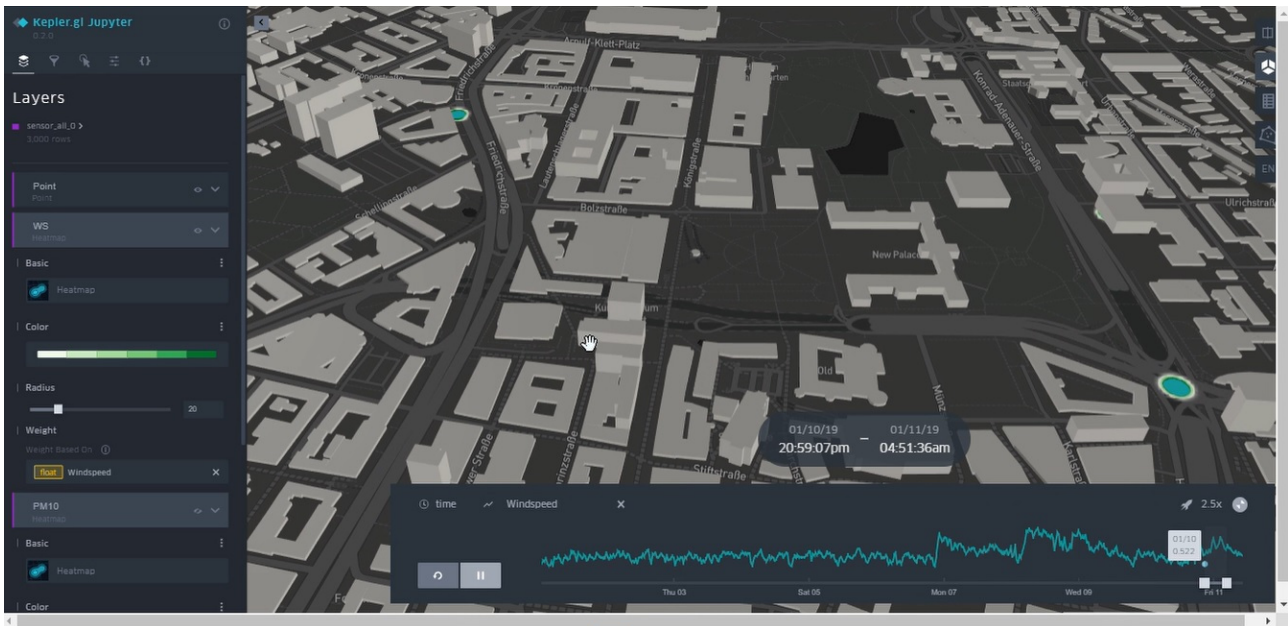


Figure 3: Sensors on map Web interface screenshot.

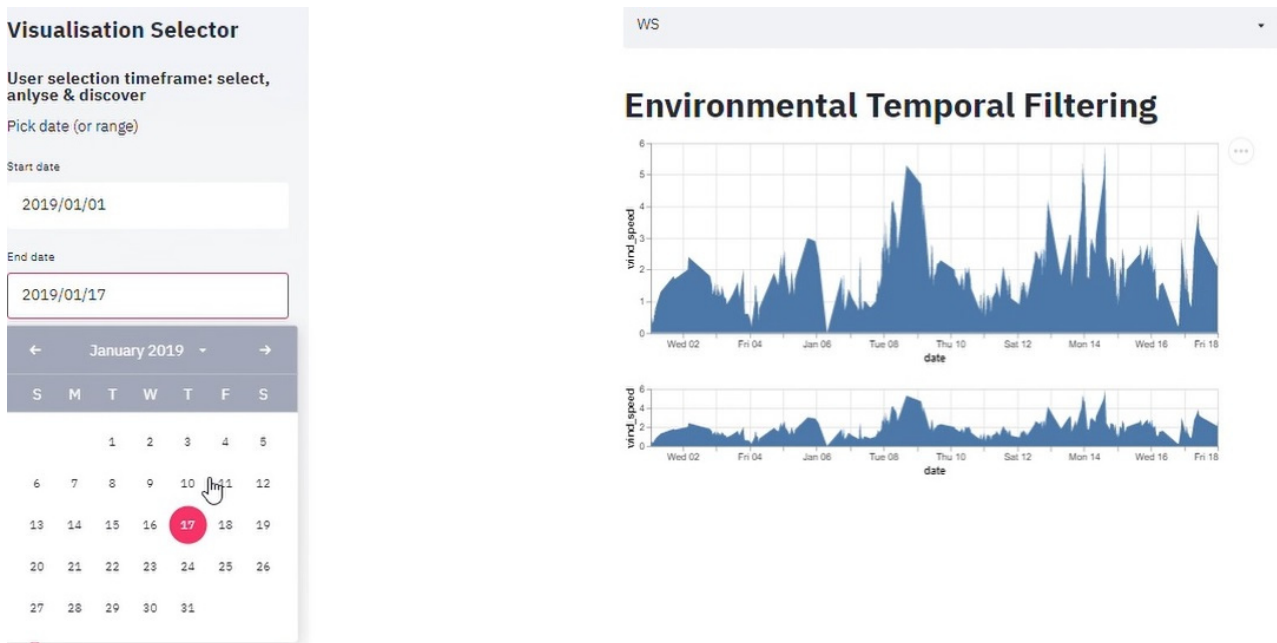


Figure 4: Show wind flow (WS) data overview in the selected temporal frame visualisation.

signed Web interface where the selected sensors' are visualised with spatial and non-spatial information over the map in the specified temporal frame. The user could select the parameters over the desired time frame and compare the patterns interactively (as shown in Figure. 3, Figure. 4, and Figure. 5). The interactive platform provides the clustering output visualisation along with sensors nature monitoring. Figure. 4, and Figure. 5 show the data overview and clustering results visualisation of designed ESVPA, respectively. Here wind speed and direction parameters are selected to analyse with the help of a stacked chart and rose plots in a chosen time frame. The output from sensors nature monitoring using Transformers Network predictions are represented using Two-Dimensional (2D) map, bar chart, and heat map as shown in Figure. 7. The developed work provides a comprehensive un-

derstanding of the relationship among the pollution parameters like PM₁₀, PM_{2.5}, and the meteorological parameters *i.e.*, wind flow and humidity.

3. DATASET

The temporal datasets of meteorological and pollution parameters are used and analysed in this study. The luftdaten selber messen provides city sensors measurements at several locations in Stuttgart, Germany. Moreover, the historical data from 2016 to 2020 from Hauptstaetter Strasse 70173 Stuttgart corner station sensor also considered. These datasets contain total eleven city centre sensors locations with wind (speed and directions), humidity along with PM_{10-2.5}, measured in a 30-minute time inter-

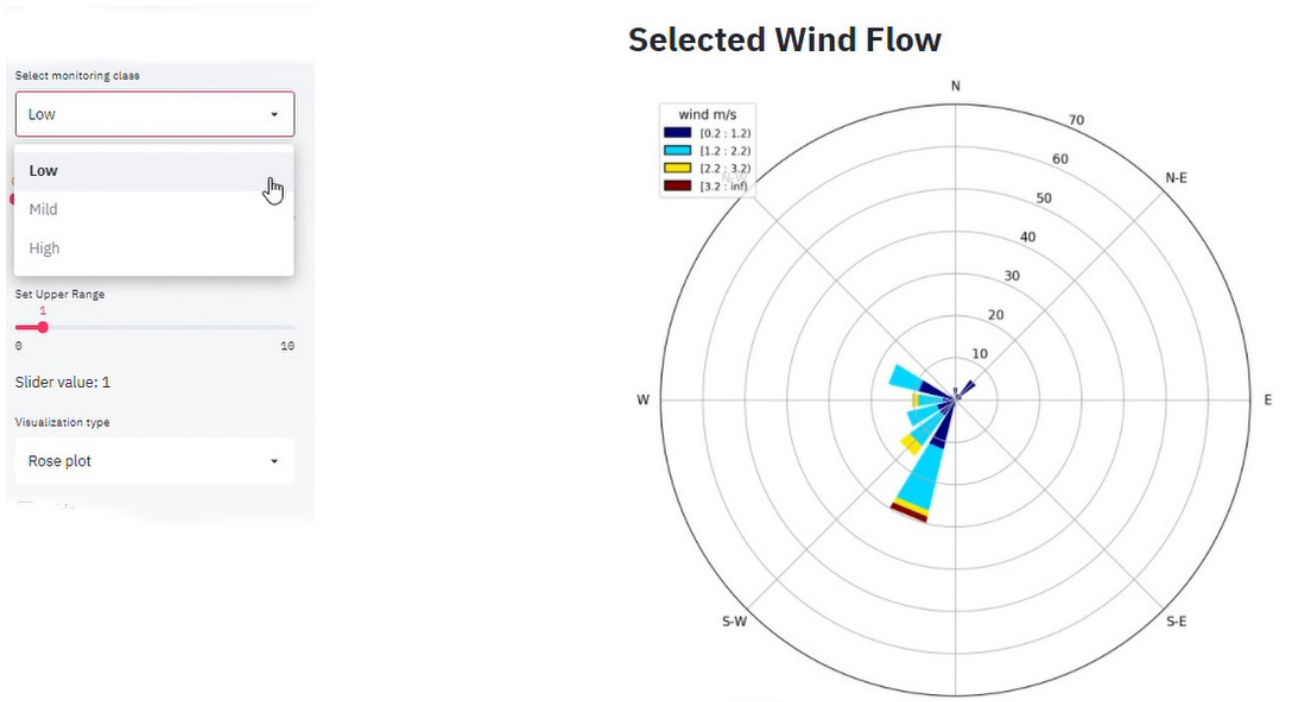


Figure 5: Rose plot generated highlighting the output of clustering sections interactive visualisation.

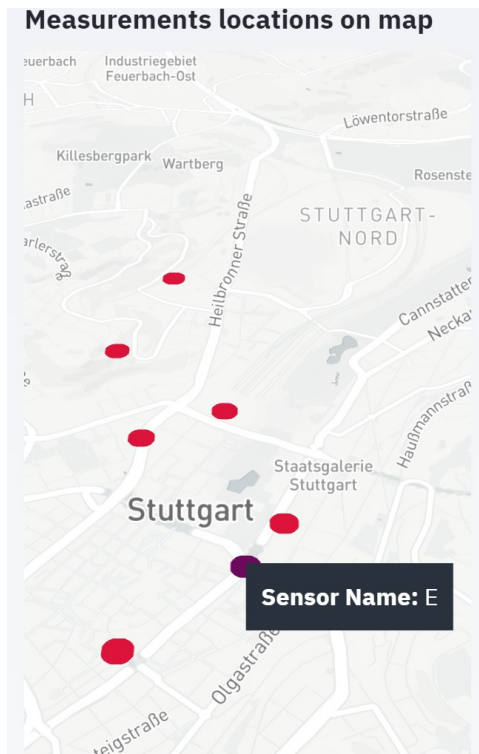


Figure 6: Displaying sensors measured locations on map.

val (Figure. 6 shows selected sensors on map). The areas dataset were organised separately into individual years for each parameter with spatial information attached, using time information with past data first, followed by current data. This helps to perform pollution and meteorological parameters temporal datasets trends and in depth analyses along with sensor's nature monitoring.

4. RESULTS

The designed algorithms and platform helps to perform in depth study of sensors measurements, and also to estimate their nature monitoring for 6hrs in future. This ESVPA is implemented as Web-based application using Altair, D3.js, kepler.gl, Streamlit, Keras library (Chollet, 2017) with TensorFlow in backend in Python and executed on Intel® Core™ i7-4770 CPU @3.40 GHz having four cores. Result in following section followed by discussion subsection analysed and validates the outcome of the proposed framework. In order to provide a more detailed comparison of the selected parameters, the visualisation of historical measurements with both spatiotemporal information attached was explored using the available interactive option of data overview in the designed platform. Figure. 2 and Figure. 3 show the historical data visualisation for the user selected time frame of PM₁₀, and wind flow on the map with the help of line charts, which helps to connect spatiotemporal information with the respective sensors measurements visually.

The platform was also used to visualise the output of the HDBSCAN clustering (as shown in Figure. 4 and Figure. 5). The unsupervised hierarchical clustering here directed for inferring the trends and inner structure of the meteorological and pollution parameters dynamically. Figure. 4 show the obtained selection for selecting clusters in the temporal data set for wind flow (WS) in selected time frame. Similar parameter analyses were conducted for rest of the parameters. Here the class value ranges of each assigned class were also displayed and compared. Moreover, the performed clustering with visualisation helped the user unboxed the complexities of datasets and their available trends in the best possible way dynamically. Furthermore, the obtained wind rose plot helped visualise wind speed and direction in a circular format in the same graph. The length of each spoke around the circle indicates the number of times (count) that the wind blows from the indicated direction. Colors along the spokes indicate classes of wind speed. Figure. 5 shows the generated wind rose plot on the selected temporal frame. Besides, each different color denoted the wind speed divided into value range boundaries at the



Figure 7: Show sensor nature monitoring example.

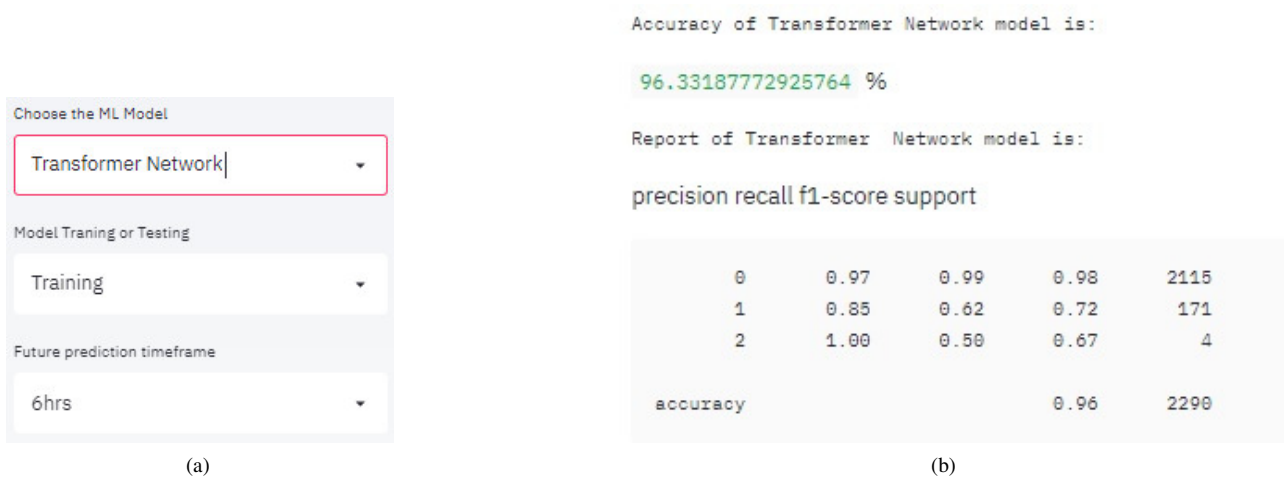


Figure 8: (a) Training interactive selection interface, and (b) model accuracy analyses visualisation.

differences (within the class assigned maximum and minimum value) with varying spoke length and direction highlighting the wind blows count from the indicated directions in this study.

Figure. 7 show the output of sensors nature monitoring using Transformers Network predictions. These options were integrated together with the selection of the desired user query. Here, the map highlights the location of the respective sensor selected, with day-wise sensors' measurements visualised with the bar chart. Moreover, the attached heat map, represented the intensity of the color, governed by the magnitude of parameter values. A similar heat map display existed for other parameters as well. Selected parameter (anyone *i.e.*, wind speed and humidity along with $PM_{10-2.5}$), having higher values (range) over the time, had been assigned a darker color in the respective heat map. In order to provide a more detailed comparison and trends analyses, user desired time frames were considered for all the parameters. Figure. 8 and Figure. 10 show the designed network achieved accuracy with the selection of the desired user query. Here, precision and recall values for predicting dominant speed for various classes for January month (of Stuttgart) are represented in Figure. 8, with reaching values for all the classes above 61% and achieved total accuracy achieved was 96.33%. Figure. 9 shows the randomly selected date for model validation. Highlighting the obtained Transformers Network visual prediction accuracy analyses with presented model success-failure (red rows). Furthermore, this has supported sensitivity analyses for calculating the success and failure of the model highlighted with color and dynamic interaction. Here networks' success was represented by

yellow color and failure with red. These color combinations were used to deliver more insides making the understanding for the user more straightforward and unboxing the complexities of ML. Thus, this platform (all together) helped to discover all the possible changes by enhancing the ability to dig in detail inside the data with accuracy for each of the considered meteorological and pollution parameters as per the user choice visually.

4.1 Discussion

The hierarchical clustering for meteorological and pollution parameters highlighted the trends at which any selected parameter was analysed in the clustering diagram, with each class assigned set lower and upper value ranges. HDBSCAN performed exploratory data analysis as it is a fast and robust algorithm that helped to work over the unsmoothed temporal meteorological and pollution parameters to return meaningful clusters. A sequential scale of color brewer for rose plot scale color map used for showing classes (low, mild and high) with the color frequency differentiates low values class from high values class. The blended progression using, typically of a multi hue, from the least to the most opaque shades with respect to value ranges lied in the clusters, represents low to high values. The 2D map view of all the selected sensors on the map, long with time based data filtering query with tool-tipping helps to easily interact and visualise all the information together in one domain Figure. 6. Each year dataset for the considered parameter over the selected time frame that joined together sooner (in clustering) are more similar to each other than those that are joined together later. The total within-

**User selected timeframe:
Predciiton Vs Original**

Start date
2020/01/01

End date
2020/01/15

Model testing: Predcited Vs Original

Graphical Demo

Figure 9: Randomly selected date for model validation: Transformers Network visual prediction accuracy analyses with presenting model success-failure (red rows).

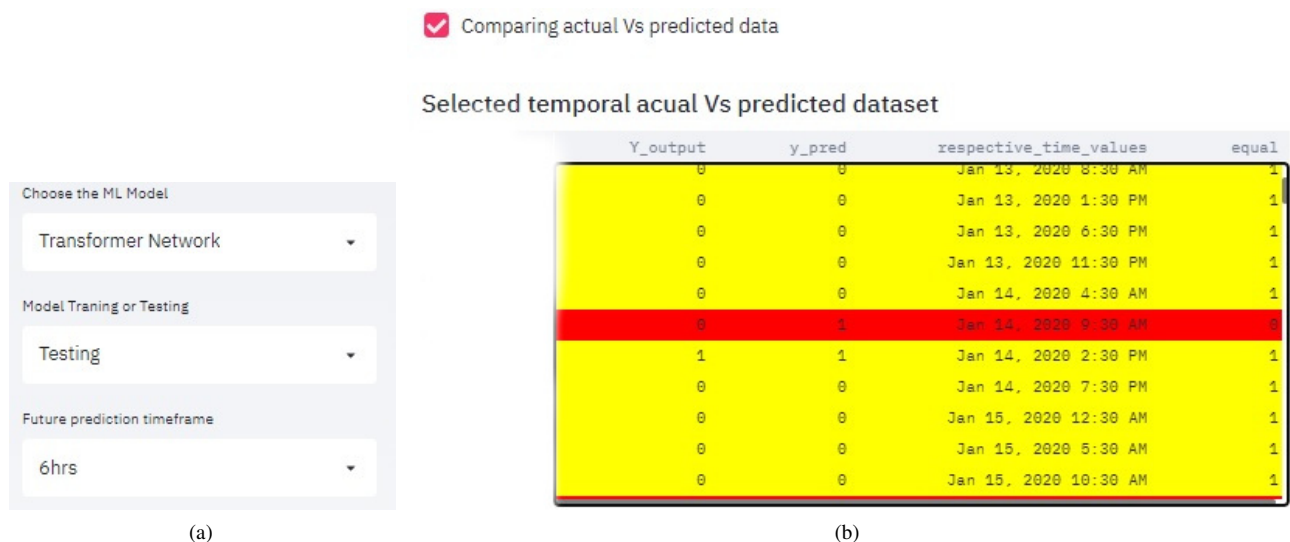


Figure 10: (a) Testing interactive selection interface, and (b) Comparative network success-failures test visualisation.

cluster variance is minimised during clustering. At each step, the paired clusters with minimum between-cluster distance are merged. As a result it is observed that in February month higher magnitude of wind flow is also observed over 2016 to 2020.

On the other hand, the Transformers Network helped to estimate sensors nature interactively. The input sample compressed of $Window_f$ consecutive values from the data with five features of $PM_{2.5}$, PM_{10} , humidity, and wind (speed and direction) provide temporal information and Transformers Network operations are able to detect trends and features. During the sample designing phase, their output classes were decided statistically using μ and σ of the total samples particular to year's data set of respective parameter (*i.e.*, anyone out of five), thereby representing the dataset better. Moreover, the total samples for a given year were divided into training and testing samples with a ratio of 7 : 3 (*i.e.*, 70% of the total of training and rest for testing). The dynamic network metrics analyses (total accuracy, precision and recall) of the Transformers Network supported with interactive visualisation helped the user verify and understand the specified with the selected parameter. Visual exploration has also been contributed to make ML more easily understandable and explainable in the sense of network inside and explainable.

Moreover, the developed ESVPA for sensors nature monitoring is used to provide interactive selections of considered meteorolo-

logical and pollution parameters to analyse the concurred pattern in the dataset, in a time frame. ESVPA is also compared with existing literature that are near to the proposed framework. Air Quality Temporal Analyser (AQTA) proposed by (Harbola et al., 2021), has provided visual analyses platform of air quality data with time but lacks sensors nature monitoring. It discovers temporal relationships among complex air quality data, on a small scale for each sensor's (individually) with missing the spatial knowledge attached. However, the developed ESVPA connect temporal, spatial and non-spatial information together visually. Further, enhancing the time series analyses using the unsupervised HDBSCAN clustering on a series of (above mentioned) parameters. Therefore, with the in-depth sensors nature understanding and trends, ML approach called Transformers Network predictor model is integrated, that takes successive time values of parameters as input with sensors' locations and predict the future dominant (highly measured) values with the location in time as the output. Thus, making ESVPA a work extension provides a big picture of sensors' nature monitoring and temporal data measurement analyses. This has helped in making the data trends analyses, and sensors nature monitoring easy, user interactive and comparable in the time domain.

5. CONCLUSION

Geovisualisation, together with visual analytics, encourages a better understanding of geospatial data by identifying trends, patterns, and contexts with making the economy, mobility, environment, people, and governance of a city smarter. In this paper, ESVPA, an interactive visualisation Web successfully designed and demonstrated for time series meteorological and pollution parameters. The temporal data analysed using the unsupervised HDBSCAN clustering on a series of these parameters. Furthermore, for sensors nature understanding and trends, Machine Learning (ML) approach called the Transformers Network predictor also integrated, which takes successive time values of parameters as input with sensors' locations and predict the future dominant (highly measured) values with location in time as the output. The interactive platform for meteorological and pollution parameters would help to plan the future with more renewable resources awareness and understanding. The designed visualisation platform (a small demonstration version) in this work and could be further improved with the ensemble of advanced visualisation approaches. The future focus for the authors would be to improve the visual analysis and utilising more advance deep learning models. Meanwhile, the devised work can help create environmental awareness among humankind and provide foreknowledge for better city planning.

ACKNOWLEDGEMENTS

The Stuttgart environmental dataset is downloaded from Department of Environmental Protection, Department of Urban Climatology (DEPUC), Stuttgart website. Moreover, more Stuttgart sensors datasets are downloaded from the luftdaten, info-Feinstaub selber messen.

REFERENCES

- Aigner, W., 2013. Interactive visualization and data analysis: visual analytics with a focus on time. *ISPRS Journal of Photogrammetry and Remote Sensing*.
- Campello, J. G. B., Moulavi, D. and Sander, J., 2013. Density-based clustering based on hierarchical density estimates. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining Proceedings of the Springer*.
- Chen, C. and Zhao, B., 2011. Atmospheric environment. Review of relationship between indoor and outdoor particles: I/O ratio, infiltration factor and penetration factor pp. 1–9.
- Chollet, F., 2017. *Deep Learning with Python*. Manning Publications Co., Greenwich, CT, USA ©2017.
- Garrett, P. and Casimiro, E., 2011. Environment science pollution. Short-term effects of fine particulate matter (PM_{2.5}) and ozone on daily mortality in Lisbon, Portugal 18, pp. 1585–92.
- Harbola, S. and Coors, V., 2018. Geo-visualisation and visual analytics for smart cities: A survey. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.
- Harbola, S. and Coors, V., 2020. Seasonality deduction platform: For pm 10, pm 2.5, no, no 2 and o 3 in relationship with wind speed and humidity. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci* p. 65–70.
- Harbola, S., Koch, S., Ertl, T. and Coors, V., 2021. Air quality temporal analyser: Interactive temporal analyses with visual predictive assessments. In: *Workshop on Visualisation in Environmental Sciences (EnvirVis)*, Vol. Accepted, p. Under Print.
- Hardin, M. and Kahn, R., 1999. *Academic press: New york. Aerosols and climate change*, 1999.
- Horvitz, E., 2007. Reflections on challenges and promises of mixed initiative interaction. In: *AI Magazine*, Vol. 53, pp. 59–67.
- Horvitz, E., 2010. Principles of mixed initiative user interfaces. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*, Vol. 53, pp. 59–67.
- Jasen, N. A. H., Fischer, P., Marra, A. C. and Cassee, F. R., 2013. Total environment. Short-term effects of PM_{2.5}, PM₁₀ and PM_{2.5-10} on daily mortality in The Netherlands 463, pp. 20–36.
- Kurcu, A., Miranda, F., Ozbay, K. and Silva, C., 2017. Data visualization tool for monitoring transit operation and performance. *IEEE Int. Conf. Model. Technol. Intell. Transp. Syst.* p. 598–603.
- Liu, S. W., Liu, M. and Zhu, J., 2017. Towards better analysis of machine learning models: A visual analytics perspective. *Vis. Informatics*. p. 598–603.
- McInnes, L. and Healy, J., 2017. Accelerated hierarchical density based clustering. In: *International Conference on Data Mining Workshops (ICDMW). Proceedings of IEEE 2017*.
- Panagiotopoulou, M. and Stratigea, A., 2017. Smart cities in the mediterranean. *ISPRS Journal of Photogrammetry and Remote Sensing*.
- Parmar, N., Vaswani, A., Uszkoreit, J., Kaiser, L., Shazeer, N., Ku, A. and Tran, D., 2018. Image transformer. In: *Proceedings of the 35 th International Conference on Machine Learning, Stockholm, Sweden*.
- Paul and Murphy, A., 2009. Data mining. Distances between Clustering, Hierarchical Clustering pp. 36–350.
- Shneiderman, B., 1996. The eye have it: a task by data type taxonomy for information visualization. In: *Proceedings of IEEE Visual Languages, College Park, Maryland*, pp. 336–343 (1996), Vol. 9, pp. 3–12.
- Sun, S., Wu, Y. and liu, S., 2013. A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology*.
- Sun, Y. and liu, S., 2016. Real-time collaborative gis: A technological review. *ISPRS Journal of Photogrammetry and Remote Sensing*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J. and Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I., 2017. Attention is all you need. In: *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA.
- Wallace, J. M. and Hobbs, P. V., 1977. *Academic press: New york. Atmospheric science, an introductory survey*.
- Wu, N., Green, B., Ben, X. and Banion, S., 2020. Deep transformer models for time series forecasting: The influenza prevalence case. In: *International Conference on Machine Learning, Vienna, Austria*.