TUM Data Innovation Lab

Munich Data Science Institute (MDSI)

Technical University of Munich

&

Bayerische Motoren Werke AG (BMW)

Final report of project:

Anomaly detection and prediction of charging station failure

Authors	Rusheel Iyer, Ludwig Graef, Ezgi Sarikayak, Davit Papikyan, Juan Carlos Climent Pardo
Mentor(s)	Dipl. Ing. Peter Huesson, M.Sc. Maya Schwaninger
Co-Mentor	M.Sc. Konstantin Riedl
Project Lead	Dr. Ricardo Acevedo Cabra (MDSI)
Supervisor	Prof. Dr. Massimo Fornasier (MDSI)

2023

Abstract

The recent global increase in the popularity of Electric Vehicles (EVs) also accelerates the demand for fast and reliable charging. In addition to ensuring the availability of charging stations, i.e. enough stations to meet the demand, ensuring that malfunctioning or broken stations are brought to the attention of both the customer and the people responsible for fixing it can go a long way in improving the customer's experience as a user. In this project, our goal is to enhance the charging experience of Bayerische Motoren Werke (BMW) EV users by guiding them away from possibly faulty charging stations. To this end we employ data-driven approaches as well as detect anomalous behavior to eventually report findings to the people accountable. First of all, we perform exploratory data analysis on historical charging session data along with charging station information provided by the BMW Group. On top of proposing a novel matching algorithm to merge the new datasets, we use both datasets, anomaly detection algorithms to identify irregular behaviors among charging stations. This allows for ranking them based on their recent performance to EV users. Hereby, we propose several approaches in how all stakeholders can be provided with the information relevant to them. The motivation is thus, that BMW EV users are able to distinguish between functioning and defect stations, save time, while enjoying a more satisfactory charging experience.

Contents

Abstract **1**

1 Introduction **3**
1.1 Anomaly detection and prediction of charging station failure 3
1.2 Problem Formulation and Approaches 3

2 Statistical Data Exploration **5**
2.1 Data Acquisition 5
2.2 Data Exploration 6
2.3 Data Pre-Processing 8
2.3.1 Data Pipeline 8
2.3.2 General Pre-processing 9
2.3.3 Sessions Pre-processing: Feature Engineering 9

3 Methodology **10**
3.1 Merging of Sessions and Stations 10
3.2 Triplet Loss - Learning an Embedding for Clustering 11
3.3 Confidence Score for Charging Stations 12
3.4 Clustering and Principal Component Analysis (PCA) 12
3.5 Trend Analysis for Failure Prediction 13
3.6 Innovation or Improvements of Algorithms or Implementations within this Project 14

4 Results **15**
4.1 Merging for Sessions and Stations 15
4.1.1 Improved classical Merging 15
4.1.2 Merging by Distributions 16
4.2 Confidence Score for Charging Stations 18
4.3 Clustering and PCA 19
4.4 Trend Analysis for Failure Prediction 20

5 Conclusions and Future Work **21**

1 Introduction

1.1 Anomaly detection and prediction of charging station failure

An Electrical Vehicle (EV) is a type of vehicle that uses one or more electric motors for propulsion, instead of an internal combustion engine. EVs convert electrical energy stored in a battery pack to mechanical energy, which is significantly more efficient compared to the conversion of fossil fuels to mechanical energy in a combustion engine. This results in lower operating costs, as well as reduced emissions of pollutants and greenhouse gases. Additionally, they can reduce the dependence on oil. Furthermore, they are quieter, smoother and more efficient than internal combustion engines. Lastly, they can also reduce the carbon footprint of transportation and help mitigate the effects of climate change. Standing with all these advantages, EVs have a promising future.

Becoming an upcoming field in the car industry, EVs have raised a number of problems that need to be addressed. One such problem is the improvement of EV user experience. Depending on various natural and anthropogenic factors, electrical charging stations degrade over time and are not able to fulfill their intended purpose. EV users visiting such faulty stations are faced with poor service (or lack thereof), which in turn calls into question the rating of EVs among users.

As one of the leading manufacturers of EVs, the BMW Group believes in the future of electrical cars. Moreover, feeling a duty towards society, BMW's goal is to reduce global emissions and meanwhile to provide the best driving experience for their customers. Having all this said, our team in cooperation with BMW, aims to enhance the charging experience of users. For this purpose we are provided with historical charging sessions information for each station in Germany in addition to static information on stations by the BMW Group. Equipped with the necessary data, our goal is to detect anomaly behaviors of charging stations, predict possible future faults, derive an availability rating for each station and recommend best-performing stations to users. We believe that this will improve the user experience, which will increase the popularity of EVs.

1.2 Problem Formulation and Approaches

Anomalies in charging stations can occur in a variety of forms, including problems with the charging equipment, issues with the power grid, and problems with the charging infrastructure itself. Some common types of anomalies include failures of the charging station, such as the charging connectors or the electrical components that convert Alternating Current (AC) power to Direct Current (DC) power. Other types of anomalies include problems with the power grid, such as voltage fluctuations or power outages, as well as issues with the charging infrastructure, such as unauthorized usage or theft.

We separate our main goal of improving the charging experience of EV users into 3 parts – ranking the stations, detecting a change in a station behavior and predicting future anomalies. With our first objective, we aim to differentiate possibly workable stations from defective ones and derive a ranking score. Next, with the second objective we intend to analyze sessions of a station in terms of the error rate observed by various detectors and identify unexpected behaviors. Finally, we want to design a data-driven system to predict future failures and guide users towards fully actionable stations.

There are several methods that are commonly used for detecting anomalies in charging stations [1–3]. Some of the popular methods include monitoring the power consumption of the charging equipment, monitoring the voltage and current of the power grid, and monitoring the usage patterns of the charging infrastructure. These methods can be used to detect problems with the charging equipment, such as failures of the charging connectors, as well as issues with the power grid, such as voltage fluctuations or power outages. Additionally, monitoring the usage patterns of the charging infrastructure can help to detect unauthorized usage or theft.

One of the main challenges in anomaly detection for charging stations is the ability to accurately detect and diagnose anomalies in real time. This is especially important in the case of power outages, as a quick response can help to minimize the disruption to the charging infrastructure. To overcome this challenge, several approaches have been proposed, such as implementing real-time monitoring and control systems that can quickly detect and diagnose problems as they occur [1–3].

Another challenge in anomaly detection for charging stations is the ability to handle the large amount of data that is generated by the charging infrastructure. This includes data from the charging equipment, such as the power consumption, voltage and current of the power grid, as well as data from the usage patterns of the charging infrastructure. To overcome this challenge, machine learning and data mining techniques are often used to analyze the data and detect anomalies.

In conclusion, anomaly detection for charging stations is a critical area of research that can help to improve the efficiency and reliability of charging infrastructure. By monitoring the power consumption of the charging equipment, the voltage and current of the power grid, and the usage patterns of the charging infrastructure, it is possible to detect and diagnose problems as they occur in real time. However, the challenge is to handle large amounts of data and provide reasonable analytics.

Another challenge will be the missing link of charging sessions to charging stations as the information on how to match stations to sessions is not available. A novel approach will be explained and discussed, in addition to an improvement to an existing solution.

The document is organized in the following manner. Section 2 provides an in-depth exploration of the charging session and charging station data. Section 3 describes the methodology and the approaches for achieving our goals. Section 4 summarizes the results and in Section 5 conclusions are derived and we discuss open points as well as propose possible directions for future work for BMW.

2 Statistical Data Exploration

This section elaborates on the data provided, the tools used to explore it, as well as initial findings on the data. Disclaimer: Specific data has not been included in the tables here due to privacy reasons, as the report may be potentially published.

2.1 Data Acquisition

The data provided by BMW consists of 2 main subsets: (i) Station Information (referred to in the following as the Station Information (SI) dataset) and (ii) Charging Session Information (referred to in the following as the Charging Session (CS) dataset).
The SI dataset, consists of information about all the existing stations that BMW EV users are able to visit when in need of a charge. This includes, for example the geolocation of the stations, provider name, operator name, and address. Figure **??** shows an extract of the SI dataset.

id	impedance	HPC	postalCode	countryCode	longitude	latitude	stationsCount	dcPower
1	AC	false	00220	FI	24	60	1	null
2	AC	false	10220	SE	25	60	1	null
3	AC	false	00270	TR	29	61	1	25
4	DC	false	02220	UK	18	70	2	null
5	AC	false	60220	FI	22	86	1	50
6	AC	true	11120	FI	38	59	1	null
7	AC	false	11321	UK	15	59	1	null
8	DC	true	79822	GE	24	55	1	28

Table 1: Extract table of the Charging Sessions data

The CS dataset, as shown below (table 2), consists of information on each charging session performed at any of these stations. The dataset stores details such as the geolocation of the car, start and end time of the charge, the amount of energy consumed (kWh), the reason the charging was ended (including possible failures), etc.

id	chargingSessionState	startTime	endTime	chargingLocation
1	FINISHED	08613472	40411005	{long: 11, lat: 52, countryCode: DE}
2	FINISHED	32122526	89948025	{long: 67, lat: 51, countryCode: TR}
3	FINISHED	07326188	42681493	{long: 35, lat: -7, countryCode: ES}
4	ERROR	17422291	83767500	{long: 4, lat: 55, countryCode: UK}
5	FINISHED	14117937	29143328	{long: 68, lat: 12, countryCode: NL}
6	UNKNOWN	01015642	90765841	{long: 7, lat: 45, countryCode: IT}
7	FINISHED	20908257	43495981	{long: 2, lat: 51, countryCode: UK}
8	FINISHED	20493896	41353594	{long: 63, lat: 18, countryCode: UK}

Table 2: Static generation of station scores over all data

Upon initial inspection of the data, a first problem becomes evident. Namely, even though geolocation information for each, sessions and stations, exists, there are no existing fields providing a direct 1-to-1 matching of sessions to stations. This means that when a charging session takes place at a hub (multiple closely-located groups of stations), it is unclear at which particular station in the group this session exactly took place. Therefore, in case of a failure, it is impossible to specify which exact station is faulty. As already mentioned, this is one of the problems that this work attempts to tackle and will be elaborated upon in Section 3.

2.2 Data Exploration

In this section, an initial data exploration is performed to get a first insight into the data to be processed. Here, the different columns of the SI and CS datasets are analyzed and plotted to see some data distributions. This subsection briefly presents a first insight.

This is especially crucial, as thanks to the data exploration the definition of a charging failure in a charging station is consolidated. If one of the following criteria is met, one could say that there is an inconsistency in the desired charging behaviour and, thus, a specific session or station has either to be further investigated or ranked differently relative to the other ones:

- Charging restrictions are not met (maxAmps, maxVolts, etc.)
- Charging goal is (not) reached
- Charging end reason
- Charging errors
- Charging start time and charging plug insert time
- Expected to real charging duration difference

Some of these characteristics are also specifically represented in some on the columns of the dataset as shown in the following images.
Figure 1 visualizes a first insight into the sessions in Europe, by depicting on the total sessions conducted in the EMEA region. It has to be noted, that a sclae indicating the number of sessions conducted is not available for technical reasons. Given that the goal is to analyze the spatial distributions of charging sessions and erroneous charging sessions, it can be argued that this is visible even without precise information about scaling.
Figure 2 in turn, shows the total number of sessions conducted in the last three years as this is the time frame over which the data provided spans. Again, it is differentiated between the total sessions conducted in the European region (left) and the ones with errors (right). The different colours represent different years. As it can be clearly observed, there is a considerable increase in the usage of EV, however, notable is that the increase in errors is not always proportional to this increase in sessions.

Figure 1: Total number of sessions on a color scale in the EMEA region. The more red an area is, the more charging sessions there are here. It seems that urban areas have significantly more charging sessions.

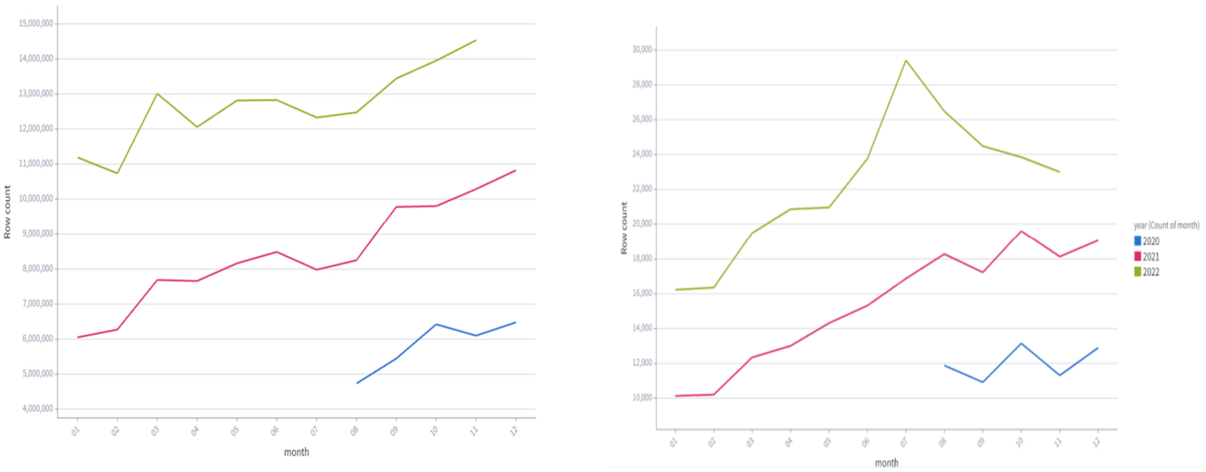

Figure 2: Total sessions (left) and sessions with errors (right) over the last three years.

Of especially importance becomes a deeper dive into the errors occurring in the sessions. Therefore, not only the different types of errors have to be considered, but also how often they appear (Figure 3). Errors 12, 13, and 31 are examples of the most relevant errors for further analysis, as they indicate whether the HV system, the charging station, or the power grid has failed, which contributes to a failed charging session.
Figure 4 shows how the data can also be used to see the different reasons behind the ending of a charging session. As highlighted, we focused on the attribute *CHARGING_STATION_FAILURE* as we want to predict when the next failure is going to happen.

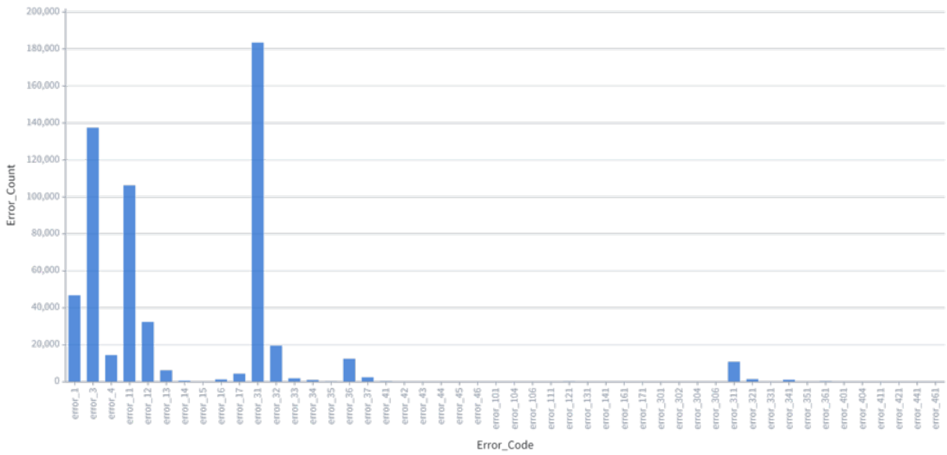

Figure 3: Distribution of business error types. Certain error types are discarded in the future analysis, as they do not contain relevant information.

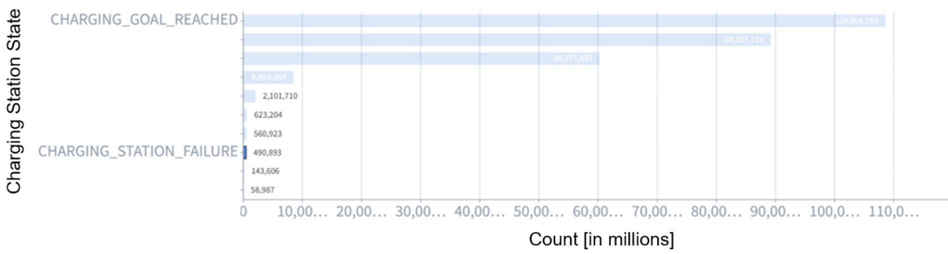

Figure 4: Next to the business errors, there is an additional attribute storing information about the charging end. As seen in the histogram, only a limited part of the charging sessions ended with a charging station failure.

2.3 Data Pre-Processing

As the data provided is already employed in production at BMW, it has already been pre-processed to a sufficient degree and is therefore relatively clean. However, the SI dataset consists, out of the box, of 82 columns and the CS dataset of 103. Multiplying, thsi by the number of session entries, the dataset is huge, and it becomes clearly, that not all columns and features are relavant fo the task at hand. For example, the operator name of the station does not provide insights into the behavior of a station. Such superfluous attributes are, therefore, identified and trimmed out of the dataset for further processing.

2.3.1 Data Pipeline

The figure below depicts the structure of the data pre-processing pipeline which shows two main branches, one for the stations on the top (marked in dark red) and one for the sessions on the middle (marked in dark purple). Each of the branches then visualizes how the pre-processing done by BMW, marked in blue, is extended by the pre-processing done by the team, marked in green. Finally, as discussed and explained in further detail in Section 3, the data is either merged (light purple box) or used for some analysis (red box).

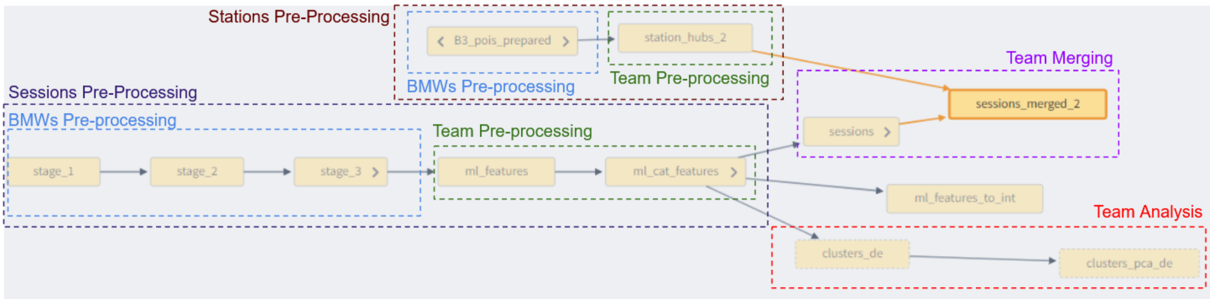

Figure 5: Pipeline for data pre-processing and posterior applications.

2.3.2 General Pre-processing

Both, in the SI and CS datasets, the data is not always stored in individual attributes per column, which results in a more complex pre-processing. Table 2 shows in the last column key-value pair-style objects, which cannot directly be fed into statistical models in later processing steps. Therefore, relevant key-value pairs have to be extracted such that the 'key' becomes a new column and the 'value' is the information it stores per row. Array-style attributes are also present in the dataset. For example, the business_error attribute stores a list of errors that might have occurred during a particular session. To deal with this, columns for each possible business error are created as a number which is set to the number of occurrences. The value is set to zero if the error did not occur during a particular session.

2.3.3 Sessions Pre-processing: Feature Engineering

Diving deeper in the middle branch of the pipeline (marked in dark purple in Figure 5), some attributes of the CS dataset have to be converted into usable data with a suitable format so that they can be used by ML algorithms. This feature engineering is done after the BMW pre-processing (marked in blue) and a list of potentially useful features is generated (marked in green).
As described in Figure 5, the feature engineering is done within two steps: First, all numerical features are processed and secondly, all categorical ones. By a simple detection of the specific data types inside of the columns, the numerical data i standardized, calculating the mean a standard deviation, and is given back to the pipeline. It has to be noted that, although numerical, the longitude and latitude attributes are excluded, as the precise values are used for merging the data with the SI dataset.
For the categorical data, specific, relevant columns are extracted and processed into a one-hot encoding format, so that they can be considered binary in the algorithms.

chargingAC	chargingDC	chargingHPC	preConditionSession	preConditionStations	sessionFinished
1	0	0	1	0	1
0	1	0	0	0	0
1	0	0	0	1	1
0	1	1	1	1	1
0	1	1	0	0	1
1	0	0	1	1	0

Table 3: One-hot encoded categorical session data after pre-processing step performed.

3 Methodology

This section addresses the various approaches and techniques used in addressing the problem and objectives introduced in Section 1. First, (i) a new algorithm to merge the SI and CS datasets is proposed. Secondly, (ii) a scoring procedure for stations based on recent behaviour is developed. As a third step, a (iii) clustering and PCA method to identify similarities or patterns in the data is introduced, and fourth (iv) a method to analyse trends for the purpose of anomaly detection is discussed.

3.1 Merging of Sessions and Stations

This section addresses the process of enhancing the merging of charging sessions to charging stations. The information at which station a charging process has taken place is not available. However, location-related information such as country, state, zip code, longitude and latitude can be found for both stations and sessions. In addition, data is provided on whether a charging station and session is AC or DC and whether a charging station can provide High Performance Charging (HPC). It is important to note, that a car can also charge at a station that is classified as HPC and does not charge at this speed. As the HPC attribute on the session side is computed, it must be guaranteed, that a session classified as non-HPC can also be merged with HPC stations. The other direction does not hold.

Previously, the distance of every charging session to every charging station was compared. The closest station is then selected. This is not scaleable with an ever-increasing number of stations and sessions. In addition, it can happen that stations and sessions with different HPC-classes can be merged together.
The approach chosen in this project provides a significant improvement. As seen in Figure 6, the merging works on subsets of the data. The sessions and stations are organized by location and impedance (AC/DC). For example, all DC sessions and stations in a specific zip code in Germany are grouped together. Additionally, there is a distinction made between HPC and non-HPC sessions and stations. Non-HPC sessions can be paired with any station, while HPC sessions can only be paired with HPC stations. The algorithm then calculates the distances between the sessions and stations in a vectorized, parallelizable manner. Finally, for each session, the closest station is selected.

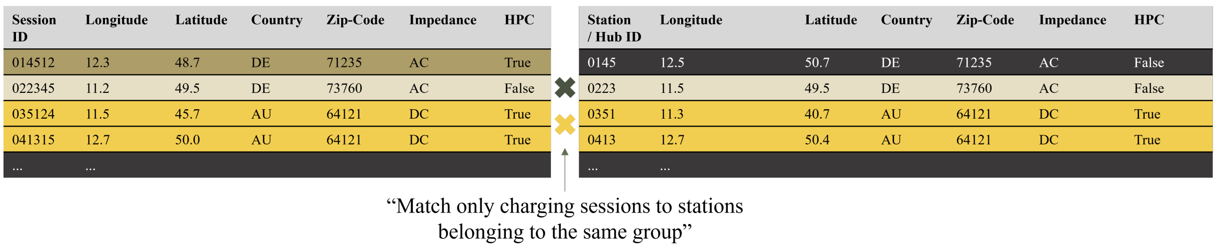

Figure 6: Schematic of Charging Station to Session Merging

3.2 Triplet Loss - Learning an Embedding for Clustering

This part of the project tests for the assumption that charging stations have a measurable difference in charging behavior. We try to learn a feature representation for a single charging session that is as similar for charging sessions on the same charging station and as dissimilar for charging stations of other charging stations.

The Triplet loss is a popular loss function used in training deep neural networks for face recognition [4] and other similar tasks. The loss function is based on the idea of anchor, positive and negative samples, where the distance from the anchor to the positive sample (i.e. charging session belonging to the same station) should have a smaller distance than the anchor to the negative sample (i.e. sessions belonging to different stations). The network should then learns to separate the features at least by a defined margin. This way, the model is trained to learn a meaningful embedding of the input data in the feature space, where semantically similar samples are close to each other, while dissimilar samples are far apart.

One common approach to clustering with the Triplet loss is to use a neural network to learn representations of the data and a cluster algorithm such as K-means to assign labels to the features [5]. The network is trained on multiple inputs, where the goal is to make the representations of equal samples (e.g. same station) closer together in the feature space, while pushing dissimilar samples further apart. The distance function can be specified. In general, the Euclidean distance is used as it ensures outliers are minimized.

The performance of the algorithm can then be measured using the Adjusted Rand Index (ARI) [6]. The score compares two clusterings. It ranges from 0 to 1. The ARI is 0 if the clustering is a random partition, while a value of 1 indicates complete equality in the cluster structure [7].

In conclusion, the Triplet loss is a versatile loss function that can be used in various machine learning tasks, including face recognition, metric learning, and image similarity learning. This project tries to apply it to distinguish charging sessions of different charging stations. The Triplet loss helps to optimize the feature representation of the input data in the feature space, such that semantically similar samples are close to each other, while dissimilar samples are far apart.

3.3 Confidence Score for Charging Stations

When EV owner wish to charge their vehicle, they often have multiple options nearby. When they arrive at a station of their choice, however, it is possible that the station has failed or does not charge as expected. This can severely affect the user's charging experience and is therefore undesirable. In order to reduce the chances of such occurrences, a 'confidence score' is generated for each station which, as the name suggests, reflects the confidence that a user can have that the station is going to charge as required.

This confidence score is built upon the recent behaviour of a station on the basis of multiple features detailed below. This means that if a station has performed poorly recently, the confidence score should in turn reduce, indicating to the user that the station might be faulty. The score is currently generated on the following features, but can easily be extended to incorporate more:

1. Total Sessions
2. Total Errors
3. Average Charging Speed

The scoring procedure goes by the following algorithm:

1. Features are aggregated on a weekly (last 7 days) basis per station
2. Feature values are standardized so that they fall between a range of 0 to 1, whereby 0 is the worst and 1 is the best possible rating
3. A weighted sum of the features is calculated according to where the weights correspond to the importance of each feature

$$s_{i,j} = \sum_{k=1}^{n} w_k \cdot x_{i,j,k} \in [0, 1] \tag{1}$$

where $s_{i,j}$ denotes the station score for station i in week j, $x_{i,j,k}$ denotes the feature value x_k ($k \in \{1, \ldots, n\}$) for station i in week j and w_k denotes the weight w for feature x_k ($k \in \{1, \ldots, n\}$), whereby $\sum_{k=0}^{n} w_k = 1$

3.4 Clustering and PCA

In order to identify any similarities in the behaviours of charging sessions, a K-means clustering [8] is carried out on all the session data. For this purpose, the number of clusters (k) is determined based on silhouette analysis [9] whereby k is limited between the range 3 to 10. This is followed by a PCA in order to visualize the clusters. The results are detailed in Section 4.3.

3.5 Trend Analysis for Failure Prediction

The popularity of EVs is rising as there is a greater demand for sustainable transportation. It is essential to be able to analyze potential failures in order to guarantee the availability and reliability for the EV charging stations. In the context of electric vehicles, charging sessions for each station includes measurements of certain attributes that are monitored over time. Another important point is that time series data usually have regular intervals, i.e. constant time intervals between data points. In contrast, intervals in our situation are irregular since the CS frequency at each station varies and is determined by the choices made by the driver. As a result, the number of sessions used varies across each window.

To detect anomalies, some variables were calculated. Simple Moving Average (SMA) is computed within a one-week windows using the number of errors for each station individually. Thus, for each window, both SMA of the number of failures and the number of sessions calculated by equation (2) and (3) respectively. Then, the error rate is derived by using two computed attributes for every window as shown by equation (4). Each station's calculations are done separately.

$$SMA_i = \frac{1}{k} \sum_{j=i}^{i+k} a_j \tag{2}$$

where SMA$_i$ denotes SMA of i^{th} window, k denotes the size of the window (1 week) and a_j denotes the number of errors in the j^{th} session inside the window.

$$C_i = \sum_{j=i-1}^{i+k-1} 1 \tag{3}$$

where C_i indicate the number of sessions happened in i^{th} window and k is the size of the window (1 week)

$$E_i = \frac{SMA_i}{C_i} \tag{4}$$

where E_i denotes the error rate of i^{th} window, SMA_i denotes the SMA of i^{th} window which is calculated with Equation (2) and C_i denotes the number of sessions happened in i^{th} window is calculated with Equation (3)

For the analysis, SMA is used to smooth out data fluctuations. In addition to SMA Gaussian smoothing [10] is used to smooth error rates.
Based on the smoothed value, quantile analysis is performed to identify anomalies, which may indicate potential failures in charging stations. At first, the average error rate for every week and every station is computed. The analysis is then based on computing the 90%-error-rate-quantile for each station individually. This serves as an upper limit for each station. If a station has a higher error rate than the 90% quantile in a given week, it is defined as an outlier.

As a second approach to identify outliers, the same procedure is also conducted for the number of sessions at charging stations. If a station has less than the 30%-number-of-sessions-quantile it is identified as an anomaly.
As a third approach for anomaly detection, the previous week is compared with the current week. If there is a sharp increase in error rates, the station is marked as outlier
To identify anomalies, Algorithm 1 is used. The results are then visualized in a BMW internal dashboard. If the quantiles are required in the algorithm, they are also displayed.

Algorithm 1 Charging stations anomaly detection

Input:
90% quantile of error rate ($quantile - 90\% - error - rate$),
30% quantile of number of sessions ($quantile - 30\% - number - of - sessions$),
lLast weeks' error rates ($LastError$),
This weeks' error rates ($NowError$),
This weeks' number of sessions ($NowCount$)
Output: Stations with anomalies

```
for each station do
    Anomaly = FALSE
    if NowError > quantile − 90% − error − rate then
        Anomaly = TRUE
    end
    if NowCount < quantile − 30% − number − of − sessions then
        Anomaly = TRUE
    end
    if NowError − 25% > LastError then
        Anomaly = TRUE
    end
end
```

This algorithm is intended to be run weekly. This ensures that stations are identified that are not already repaired at the time of the analysis.

3.6 Innovation or Improvements of Algorithms or Implementations within this Project

This project includes several different sub-projects, each of which seeks to either improve existing algorithms or develop new approaches. The following projects are continued work of existing algorithms:

- Data cleaning and normalization,
- Data analysis and visualization,
- Clustering to detect hidden structures in the data,
- Trend analysis based on existing characteristics (number of sessions, error rates),
- Merging by location and charging characteristics.

In addition to the approaches already mentioned, several others were tested. Since these did not provide any further insights, they are not mentioned here. Along with the modified algorithms, several algorithms were not implemented before the project and could be continued in the future:

- Station scoring based on nearby charging stations,
- Merging charging stations and charging sessions by charging characteristics

This chapter gave an introduction on the methodology and a base understanding of the underlying algorithms. The next chapter will provide details on the aforementioned implementations.

4 Results

This project resulted in some insights regarding the analysis of the charging station and session data. The key results are analyzed in the following section. Furthermore, their limitations and the further outlook is discussed.

4.1 Merging for Sessions and Stations

One task of this project is the improvement of the already existing merging procedure of charging sessions and stations. It is possible to match charging stations and sessions based on their location properties as well as their power characteristics (e.g. impedance and HPC/non-HPC). Station hubs are defined as multiple stations with the same properties, that are not distinguishable by the location or power characteristics.
After matching charging sessions to hubs it might be possible to distinguish different charging stations based on their charging characteristics. It is crucial to understand that sessions cannot be traced to a particular station since they are assigned a dummy identifier. Consequently, it is impossible to identify a single station. However, errors can be assigned to a station through the dummy label. For instance, if one hundred errors are observed at a hub with five charging stations, it becomes possible to assign the errors to individual stations instead of attributing them to all five stations combined. Nonetheless, it is still not possible to clearly match the stations to the dummy labels.
The following section first analyzes an improved version of the standard matching. Subsequently, an approach to distinguish charging sessions on charging hubs based on the Triplet loss (Section 3.2) is discussed.

4.1.1 Improved classical Merging

The matching based on latitude and longitude introduces certain limitations. An increasing number of charging stations and sessions is added to the systems. At the moment, the algorithm iterates over all charging sessions and checks which is the closest charging station. In terms of computational complexity, this scales at

$$O(\#sessions * \#stations)$$

Simply speaking, the approach tested in this project is different in that sense, that it does not try to match charging stations in Germany to those located in France. Based on location features, such as the country-code and zip-code, that are available for charging stations and sessions, the algorithm groups matching candidates. Furthermore, subgroups based on power characteristics are formed. In addition, the distance computation is improved, as it relies on a vectorized function that reduces compute time. The overall complexity should then be decreased to a complexity of

$$O(\#sessions * log(\#stations))$$

The logarithmic scaling arises from the fact that every additional station does not linearly increase the number of computations for all sessions but only in a single sub-group.
In Table 4 a comparison of the performance of the previous to the current approach is illustrated.

Algorithm	Compute Time
Current Approach	55:43 minutes
Our Approach	5:59 minutes

Table 4: Results of the different approaches used for matching charging stations and sessions.

The speed improvement is significant. Nonetheless, it needs to be mentioned that the number of matches is reduced. The matching requires a zip-code that is not always available on the charging station side. Charging stations without a zip-code are discarded and cannot be matched. In an additional step, it would be possible to also match charging stations and sessions without zip-code. This would then increase the size of the groups and reduce the overall performance.

4.1.2 Merging by Distributions

In some cases it is not possible to match a charging session with a station. If stations cannot be differentiated by their location or power characteristics they are grouped as hubs. In order to differentiate different stations in a single hub, this project analyzes the possibility to assign sessions to stations by analyzing charging characteristics (e.g. charging time, charging power, maximum charging speed, etc.).
The algorithm used in this scenario is based on the Triplet loss as discussed in Section 3.2. The procedure works as follows:

1. Normalize charging characteristics of sessions.

2. Identify charging hubs with a single charging station.

3. Select charging stations that are the same model and have the same power.

4. Group these charging stations and their respective charging sessions together.

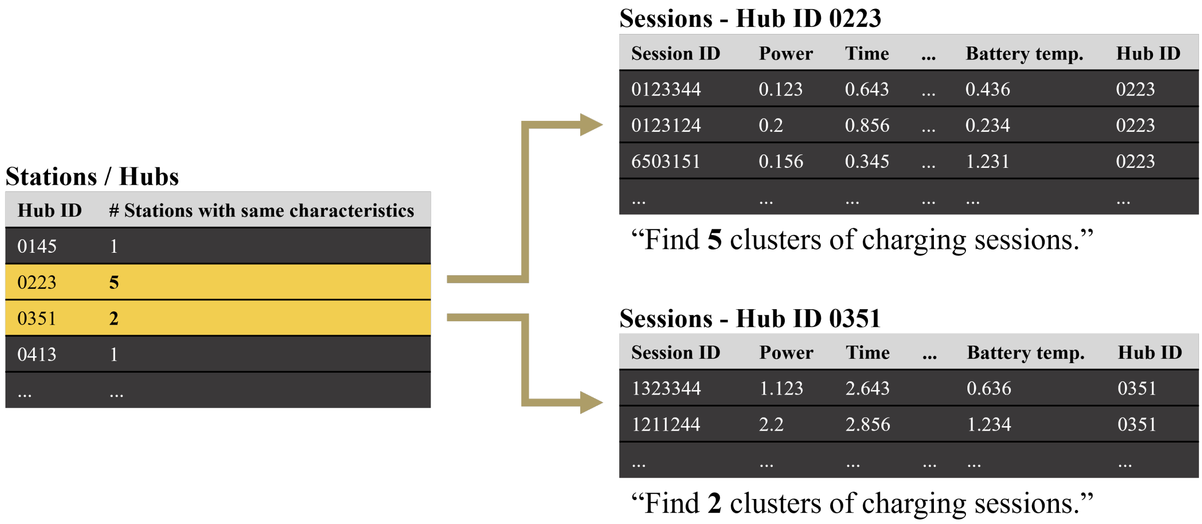

Figure 7: Schematic of differentiating different charging stations. Charging stations with more than one station are not distinguishable by their location and power properties. Cluster charging stations according to their charging session characteristics.

5. Train a Neural Network using the Triplet loss to generate features that maximize the distance between sessions from the same station and sessions from different stations.

6. Apply K-means clustering on the resulting features to group similar sessions.

Once the algorithm is tested on different unseen charging station types, it can be applied to charging hubs with multiple stations.
The performance of the model is then evaluated using the previously introduced ARI (see Section 3.2). In this case, the ARI compares the result of the K-means clustering to the ground truth station identifiers. As the algorithm only considers charging hubs with a single station, the ground truth labels are still available.
In order to make the results comparable, a second feature transformation is tested. The PCA is reducing dimensionality by maximizing the explained variance of the remaining components. The output of the Neural Network and the PCA transformation is a ten dimensional feature space.

Model	ARI (test set)	ARI (unseen station type)
PCA & K-means clustering	0.059	0.018
Triplet loss & K-means clustering	0.203	**0.188**

Table 5: Results of using PCA compared to Triplet loss. Overall performance is low, but learned features at one station type can be transferred to another station type.

In Table 5 results are shown. Interestingly, the clustering works significantly better when training a Neural Network model over running a linear PCA. Nonetheless, the performance

is not sufficient. An ARI under 0.8 is probably not satisfactory enough to conduct further analyses based on this clustering. It is perhaps possible to improve performance by considering more features that capture more station specific characteristics. It is also important to note, that the training was only performed on a single station type. Training on multiple station types could influence the performance. It is interesting to observe that the performance is almost similar in differentiating stations from the same station type (test set) to completely unseen stations. This suggests that the Triplet loss extracts some charging station specific characteristics from the features.
In general, the extracted features from the Triplet loss are not expressive enough to perform station clustering. Nonetheless, they might provide sufficient signal to correlate them to other charging properties, such as the error rate.

4.2 Confidence Score for Charging Stations

Two possible approaches present themselves when implementing the confidence score:

1. Calculate the score for all stations and every week across the entire dataset
2. Get user data (e.g., car location) and calculate the score exclusively for each station in a certain radius and based on it's behaviour only in the most recent week

Hereby, the first option is far more resource intensive but makes the data readily available when it is needed.

Date	Station	Error_Rate	Charge_Speed	Score
13-12-2022	iComp_0843	0.40	0.18	0.67
19-07-2022	iComp_0843	0.07	0.18	0.87
13-12-2022	195472	0	0.18	0.96
19-07-2022	195472	0.09	0.18	0.90

Table 6: Static generation of station scores over all data

Table 6 displays an extract of what this would look like in production. Here, sample values can be seen for each of the features per station per week. Important to note is the impact of the error rate on that station score, due to its high weightage. As station iComp_0843 has an error rate of 0.40 in the week 13-12-2022 - this means 40% of the sessions returned an error - and therefore has a low station score of 0.67. However, a strong decrease in the error rate increases its score by 20% in the following week which means that the customer can be more confident in having a successful charge session if they choose to visit this station.

The second option, while less resource intensive, requires a higher inference time as the scores must be generate when the user requests it. Furthermore, communication between the user and the data server must be established. The latter, however, also could potentially allow the user to make additional requirements on the station-level, such as whether

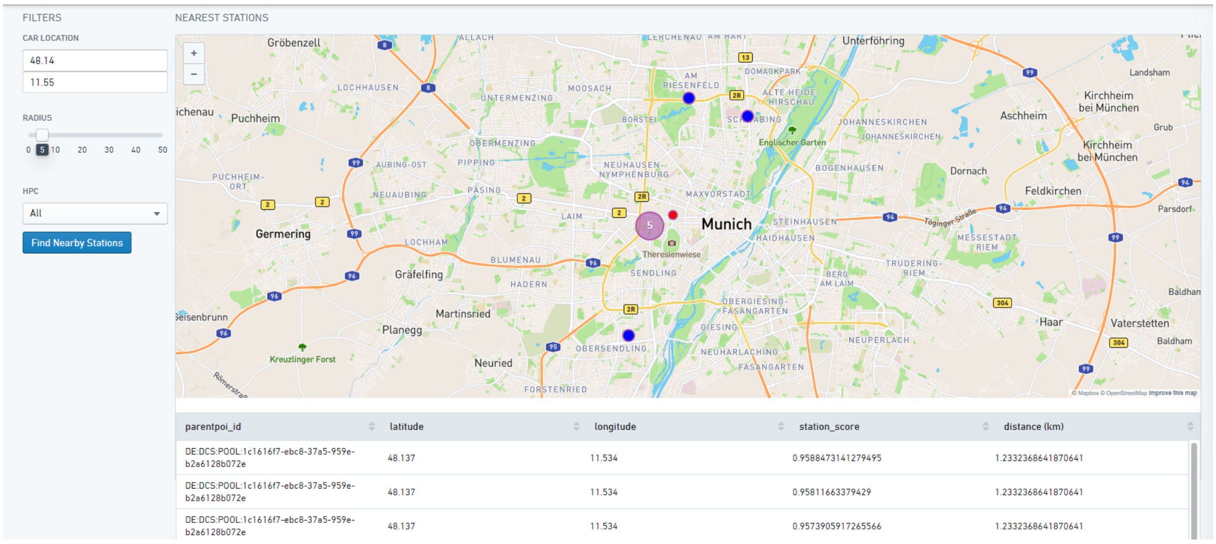

Figure 8: Application for visualization and dynamic calculation of station scores based on customer information

they require an HPC charging station and so on.

Figure 8 illustrates how the results of this approach could be displayed to the user. On the left-hand side, a customer's attributes and filters are shown, such as car location, radius for the station search and whether or not they would only like to visit an HPC station. In the map, the red dot represents the customer's location, and the blue dots the stations in the selected radius. The table below the map, shows the location, scores and distances of each of the stations from the customer. While the customer is shown the station with the best score first, they are free to decide which of these stations they would like to visit based on their continued journey and preferences.

4.3 Clustering and PCA

As explained in Section 3.4, K-means clustering is applied on the session data. Due to initial resource limitations, in this work, the analysis has only been limited to sessions taken place on or after 01-10-2021 in Germany. However, this can be easily extended to the rest of the data. The silhouette analysis determines $k = 7$ as the optimal number of clusters with a silhouette score of 0.92. Silhouette scores fall within a range of $[-1, 1]$ where a score closer to 1 reflects densely-clustered and well-separated clusters [11]. The score for session clustering therefore reveals very well separated clusters. This is illustrated in Figure 9 where the sessions (limited to the first 500,000 sessions) are plotted against the first two principle components from the PCA.

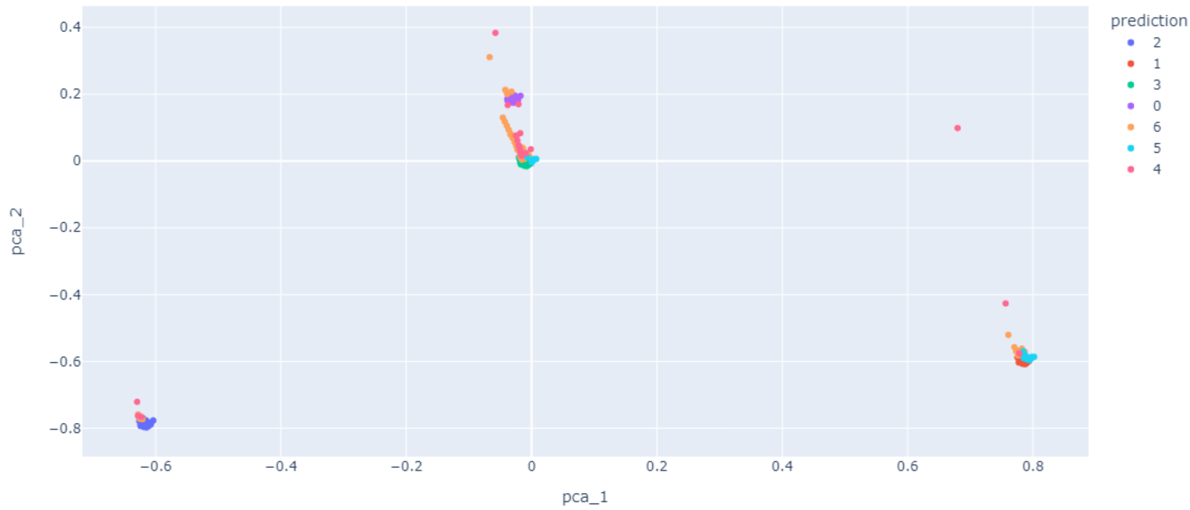

Figure 9: Clustering results based on PCA

An extensive analysis and interpretation of the clusters could not be carried out in the scope of this project and is therefore left to future work. However, a noteworthy result from initial inspection of the clusters shows that the majority (over 99%) of sessions that report a charging station or High Voltage (HV) system failure, charging system or external error belong to cluster 3. This would suggest that future sessions that are assigned to cluster 3 can be flagged early as sessions with a high probability of having failed.

4.4 Trend Analysis for Failure Prediction

From a maintenance perspective it is interesting to identify stations with a significant share of faulty charging sessions. If a station is identified, it is possible to trigger other actions that could either prevent BMW customers to access this station, or contact partner companies that inspect the station.

This project introduces several techniques to identify defect stations as shared in methodology part. A visual prototype has been created, that can be used to analyze stations on a weekly bases.

Station/Hub ID	Error Rate-Upper Quantile	Error Rate-Last Week	Error Rate-Current Week
0145	0.3	0.2	0.1
0223	0.1	0.4	**0.55**
0351	0.25	0.3	0.27
0413	0.03	0.02	0.03

Table 7: Static generation of station scores over all data

In this table, we used dummy data. Error rates or drops in the number of sessions, which can indicate a potential problem with the charging system as shown in Section 3.5. The important key points are summarized below.

- Lower quantile for the number of charging session is 30% for each station. If the number of sessions are lower than the lower quantile, it is defined as an anomaly.
- Upper quantile for the number of charging session is 90% for each station. If the error rate is higher than the upper quantile of the station specific error rate, we define it as an outlier.
- If the percentage of sessions that have an error is 25% higher than last week's error rate, this is another indication of an anomaly.

The mentioned thresholds can be defined as desired. Lowering the requirements would result in a higher recall, whereas increasing the requirements increases precision. It depends on the use case what is desired.

In general, identifying such anomalies, as shown in Section4.4, can help BMW to take action to solve problems in the stations and increase the customer satisfaction. The extracted data can then be visualized. A prototype dashboard is created, but cannot be shared in this report for privacy reasons. The prototype could serve as the basis for a charging station detector.

5 Conclusions and Future Work

Improving consumer satisfaction requires a focus on the performance of electric charging stations. This project aims to make a modest contribution towards that goal. With the assistance of this project BMW has obtained a solution for a faster merging algorithm for sessions to stations. We create an algorithm for calculating the confidence score of different charging stations based on variables such as total errors, total sessions, and charging speed. We apply a clustering method to detect similarities in the behavior of the loading processes. In addition, from trend analysis we can find the stations with abnormal behavior with a large increase in error rates and anomalous decrease on the charging sessions. Finally, a possible way to extract station-specific features using the Triplet loss is presented.

The following summarizes potential areas for future work building upon the contributions of this report:

- Customer Interface Design: Create a user-friendly interface for the station score. Customers should be able to simply and quickly access and view the scores of charging station using this interface. While we proposed a prototype in this report, further work needs to be done in refining it and on how it can be implemented on the customer side.
- Study of Clusters and BMW: Conduct a more in-depth analysis of the clusters to better understand any relationships that could be discovered via the clustering.

- Time Series Forecasting: Usage of more advanced time series models such as Hidden Markov Models (HMM), Long Short-Term Memory (LSTM), and Transformers to increase the accuracy and resilience of station score predictions. These models can aid in the detection of complicated patterns and trends in charging station data.
- External features: Usage of external features influencing the lifespan of charging stations and the charging procedure itself. Such features include the temperature of the environment, humidity, precipitation, pressure, etc. We strongly believe that these attributes have the potential to positively influence the construction of the charging score.
- Charging station features: Leverage the features generated by the Triplet loss to identify station behavior, and integrate with additional algorithms to forecast station failures and other relevant outcomes.

This project serves as baseline for further analysis. Several improvements and novel approaches have been discussed. Others can base their research on these insights for more advanced algorithms and different approaches.

References

[1] Pere Izquierdo Gómez, Alberto Barragan Moreno, Jun Lin, and Tomislav Dragičević. Data-driven thermal modelling for anomaly detection in electric vehicle charging stations. In *2022 IEEE Transportation Electrification Conference & Expo (ITEC)*, pages 1005–1010. IEEE, 2022.

[2] Jesus Cumplido, Cristina Alcaraz, and Javier Lopez. Collaborative anomaly detection system forÂ charging stations. In Vijayalakshmi Atluri, Roberto Di Pietro, Christian D. Jensen, and Weizhi Meng, editors, *Computer Security – ESORICS 2022*, pages 716–736, Cham, 2022. Springer Nature Switzerland.

[3] Yue Yang and Jiarou Li. Electric vehicle charging anomaly detection method based on multivariate gaussian distribution model. In *Proceedings of the Asia Conference on Electrical, Power and Computer Engineering*, EPCE '22, New York, NY, USA, 2022. Association for Computing Machinery.

[4] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[5] Kalun Ho, Janis Keuper, Franz-Josef Pfreundt, and Margret Keuper. Learning embeddings for image clustering: An empirical study of triplet loss approaches. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 87–94. IEEE, 2021.

[6] Jorge M Santos and Mark Embrechts. On the use of the adjusted rand index as a metric for evaluating supervised classification. In *Artificial Neural Networks–ICANN 2009: 19th International Conference, Limassol, Cyprus, September 14-17, 2009, Proceedings, Part II 19*, pages 175–184. Springer, 2009.

[7] Yun Yang. *Temporal data mining via unsupervised ensemble learning*. Elsevier, 2016.

[8] Aristidis Likas, Nikos Vlassis, and Jakob J. Verbeek. The global k-means clustering algorithm. *Pattern Recognition*, 36(2):451–461, 2003. Biometrics.

[9] Meshal Shutaywi and Nezamoddin N. Kachouie. Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6), 2021.

[10] Alle Meije Wink and Jos BTM Roerdink. Denoising functional mr images: a comparison of wavelet denoising and gaussian smoothing. *IEEE transactions on medical imaging*, 23(3):374–387, 2004.

[11] Ketan Rajshekhar Shahapure and Charles Nicholas. Cluster quality analysis using silhouette score. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 747–748, 2020.