**Machine Learning Engineer Nanodegree**

**Capstone Report (LANL Earthquake Prediction from Kaggle Competition)**

Tobias Steidle

([tobias.steidle@softwaredev.de](mailto:tobias.steidle@softwaredev.de))

April xxth, 2019

**Report**

# I. Definition

### Projekt Overview

In diesem Projekt geht es um die Vorhersage des Zeitpunkts wann ein Erdbeben eintritt. Dabei soll basierend auf seismischen Daten mittels eines Machine Learning Verfahrens ermittelt werden wann das Erdbeben auftreten wird bzw. wieviel Zeit noch bis zum Eintritt des Erdbebens vergehen wird.

Das Projekt ist als Kaggle Challenge angelegt. D.h. die Trainings – und Testdaten werden auf der Kaggle Plattform zur Verfügung gestellt. Und es ist ein Wettbewerb um den niedrigsten „Mean Absolut Error“ auf die Testdaten zu erreichen. Wenn das Problem gelöst wird kann das trainierte Model auf reale Daten skaliert werden. Damit haben die Forscher das Potential die Vorhersagen von Erdbebengefahren zu verbessern und damit Leben und Milliarden von Dollar an Infrastruktur retten können.

### Problem Statement

Das Ziel in diesem Projekt ist einen möglichst geringen MAE (Mean Absolute Error) auf dem Testdatensatz zu erreichen. Ein möglichst kleiner MAE bedeutet in diesem Fall das der vorhergesagte Zeitraum bis zum eintreten des Erdbebens nahe an den realen Zeiträumen liegt den Testdaten zugrunde liegen.

Um dieses Problem zu lösen sind folgende Schritte notwendig:  
1. Download der Daten von Kaggle  
2. Visualisieren der Daten zum besseren Verständnis  
3. Preprocessing der Daten  
4. Trainieren eines passenden Regressors  
5. Schreiben der Submission File für die Kaggle Competition

Die Submission file wird anschießend zum festlegen des Scores an Kaggle übertragen.

### Metrics

The used metric is the Mean Absolute Error. The Mean Absolute Error measures the average magnitude of the errors in a set of predictions, without considering their direction. It’s the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight.

= Number of predicted values  
 = real value  
 = predicted value

For each record oft he test dataset the prediction error is calculated. Convert each error to a positive figure by taking the absolute value for each error. Finally, calculate the mean value for all recorded absolute errors.

Da es sich hier um eine Kaggle Competition handelt ist die Metrik in diesem Fall vorgegeben.

# II. Analysis

### Data Exploration

#### The data sets are freely available for download on the Kaggle competition website.

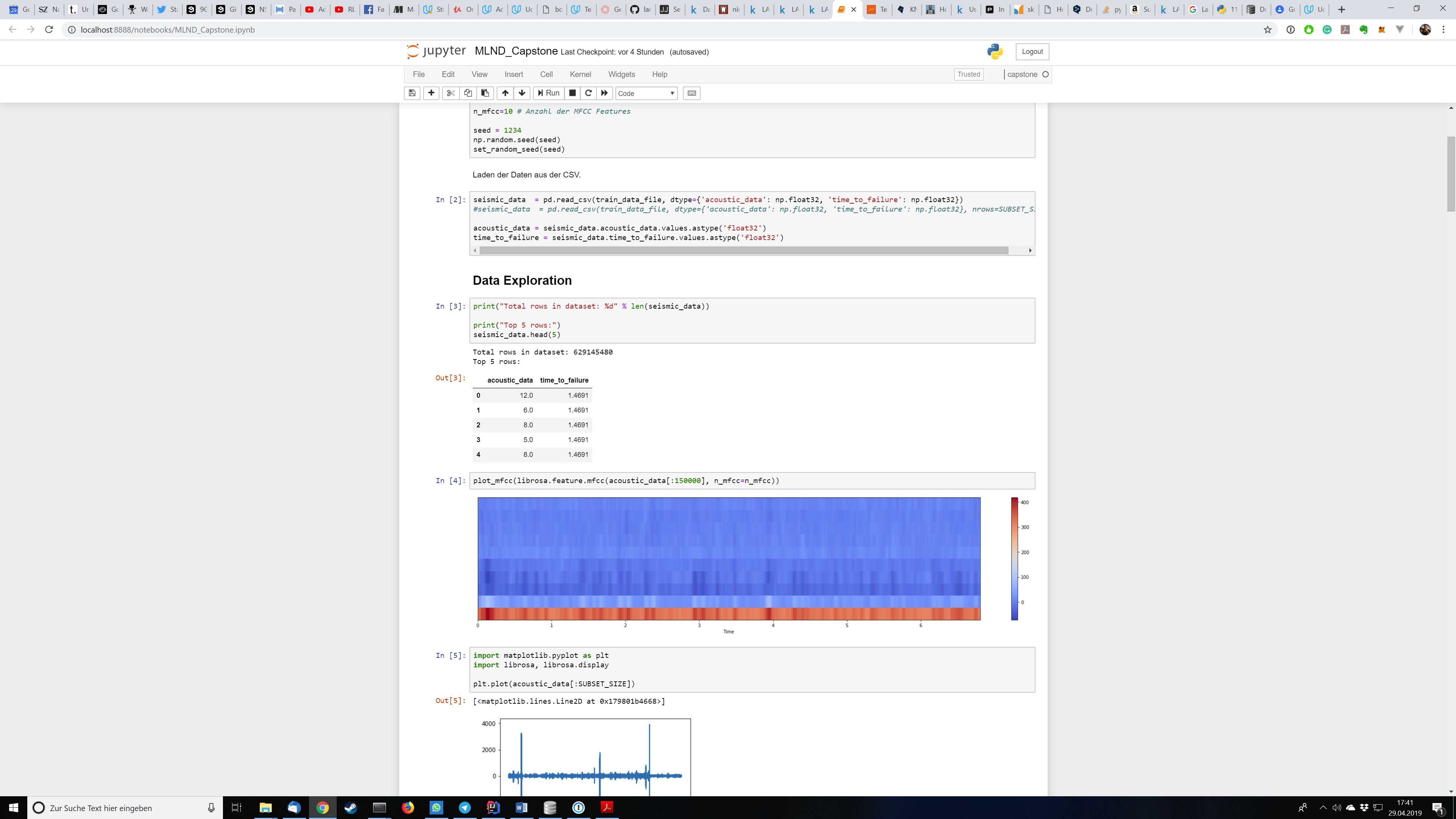
#### The data consists of a train.csv file containing a single, continuous training segment of experimental data. And test data in the test.zip file.

#### Input Data fields

* acoustic\_data - the seismic signal [int16]
* time\_to\_failure - the time (in seconds) until the next laboratory earthquake [float64]
* Total rows in this train dataset: 629145480.

#### Test data set

* test.zip containing many small segments of test data.
* test dataset that is split into 2624 files , each one containing 150,000 acoustic\_data points

  
Auszug aus den Trainingsdaten

Was auf einen Blick auf den Datenauszug nicht ersichtlich ist, ist das der „time\_to\_failure“ in diesen 5 Zeilen zwar identisch aussieht, es jedoch nicht ist. Es sind im realen Datensatz mehrere Stellen hinter dem Komma vorhanden. Die „time\_to\_failure“ ist also in Wahrheit absteigend bis zum Auftreten des Erdbebens. Nach dem Erdbeben wird die „time\_to\_failure“ wieder hochgesetzt. D.h. in den Trainingsdaten sind mehrere Segmente (Start der Messung bis Eintritt des Erdbebens) hintereinander zusammengefasst.  
In den Testdaten beinhaltet jede Datei ein komplettes Segment.

### Exploratory Visualization

he data sets are freely

*### Exploratory Visualization*In this section, you will need to provide some form of visualization that summarizes or extracts a relevant characteristic or feature about the data. The visualization should adequately support the data being used. Discuss why this visualization was chosen and how it is relevant. Questions to ask yourself when writing this section:  
- \_Have you visualized a relevant characteristic or feature about the dataset or input data?\_  
- \_Is the visualization thoroughly analyzed and discussed?\_  
- \_If a plot is provided, are the axes, title, and datum clearly defined?\_

In diesem Abschnitt müssen Sie eine Form der Visualisierung bereitstellen, die ein relevantes Merkmal oder Merkmal über die Daten zusammenfasst oder extrahiert. Die Visualisierung sollte die verwendeten Daten angemessen unterstützen. Diskutieren Sie, warum diese Visualisierung gewählt wurde und wie sie relevant ist. Fragen, die Sie sich beim Schreiben dieses Abschnitts stellen sollten:

- Haben Sie ein relevantes Merkmal oder Merkmal über den Datensatz oder die Eingangsdaten visualisiert?

- Ist die Visualisierung gründlich analysiert und diskutiert?\_

- Wenn ein Plot erstellt wird, sind die Achsen, der Titel und der Bezugspunkt klar definiert?\_

*### Algorithms and Techniques*In this section, you will need to discuss the algorithms and techniques you intend to use for solving the problem. You should justify the use of each one based on the characteristics of the problem and the problem domain. Questions to ask yourself when writing this section:  
- \_Are the algorithms you will use, including any default variables/parameters in the project clearly defined?\_  
- \_Are the techniques to be used thoroughly discussed and justified?\_  
- \_Is it made clear how the input data or datasets will be handled by the algorithms and techniques chosen?\_

In diesem Abschnitt müssen Sie die Algorithmen und Techniken besprechen, die Sie zur Lösung des Problems verwenden möchten. Sie sollten die Verwendung jedes einzelnen basierend auf den Merkmalen des Problems und der Problemdomäne begründen. Fragen, die Sie sich beim Schreiben dieses Abschnitts stellen sollten:

- Sind die Algorithmen, die Sie verwenden werden, einschließlich aller Standardvariablen/Parameter im Projekt klar definiert?\_

- Sollen die Techniken gründlich diskutiert und begründet werden?\_

- Wurde klargestellt, wie die Eingangsdaten oder Datensätze von den gewählten Algorithmen und Techniken behandelt werden?

*### Benchmark*In this section, you will need to provide a clearly defined benchmark result or threshold for comparing across performances obtained by your solution. The reasoning behind the benchmark (in the case where it is not an established result) should be discussed. Questions to ask yourself when writing this section:  
- \_Has some result or value been provided that acts as a benchmark for measuring performance?\_  
- \_Is it clear how this result or value was obtained (whether by data or by hypothesis)?\_

In diesem Abschnitt müssen Sie ein klar definiertes Benchmark-Ergebnis oder einen Schwellenwert für den Vergleich der von Ihrer Lösung erzielten Leistungen angeben. Die Argumentation hinter dem Benchmark (falls es sich nicht um ein bestimmtes Ergebnis handelt) sollte diskutiert werden. Fragen, die Sie sich beim Schreiben dieses Abschnitts stellen sollten:

- Wurde ein Ergebnis oder ein Wert geliefert, der als Maßstab für die Leistungsmessung dient?\_

- Ist es klar, wie dieses Ergebnis oder dieser Wert erhalten wurde (ob durch Daten oder durch Hypothese)?\_  
  
*## III. Methodology*\_(approx. 3-5 pages)\_  
  
*### Data Preprocessing*In this section, all of your preprocessing steps will need to be clearly documented, if any were necessary. From the previous section, any of the abnormalities or characteristics that you identified about the dataset will be addressed and corrected here. Questions to ask yourself when writing this section:  
- \_If the algorithms chosen require preprocessing steps like feature selection or feature transformations, have they been properly documented?\_  
- \_Based on the \*\*Data Exploration\*\* section, if there were abnormalities or characteristics that needed to be addressed, have they been properly corrected?\_  
- \_If no preprocessing is needed, has it been made clear why?\_

In diesem Abschnitt müssen alle Ihre Vorverarbeitungsschritte klar dokumentiert werden, falls erforderlich. Aus dem vorherigen Abschnitt werden alle Anomalien oder Merkmale, die Sie in Bezug auf den Datensatz identifiziert haben, hier behandelt und korrigiert. Fragen, die Sie sich beim Schreiben dieses Abschnitts stellen sollten:

- Wenn die gewählten Algorithmen Vorverarbeitungsschritte wie Featureauswahl oder Featuretransformationen erfordern, sind sie dann ordnungsgemäß dokumentiert?\_

- Basierend auf dem Abschnitt \*\*Data Exploration\*\*, wurden Anomalien oder Merkmale, die behoben werden mussten, ordnungsgemäß korrigiert?\_

- Wenn keine Vorverarbeitung erforderlich ist, wurde dann klargestellt, warum?\_

*### Implementation*In this section, the process for which metrics, algorithms, and techniques that you implemented for the given data will need to be clearly documented. It should be abundantly clear how the implementation was carried out, and discussion should be made regarding any complications that occurred during this process. Questions to ask yourself when writing this section:  
- \_Is it made clear how the algorithms and techniques were implemented with the given datasets or input data?\_  
- \_Were there any complications with the original metrics or techniques that required changing prior to acquiring a solution?\_  
- \_Was there any part of the coding process (e.g., writing complicated functions) that should be documented?\_

In diesem Abschnitt muss der Prozess, für den Sie Metriken, Algorithmen und Techniken für die angegebenen Daten implementiert haben, klar dokumentiert werden. Es sollte völlig klar sein, wie die Umsetzung durchgeführt wurde, und es sollte über alle Komplikationen diskutiert werden, die während dieses Prozesses aufgetreten sind. Fragen, die Sie sich beim Schreiben dieses Abschnitts stellen sollten:

- Wurde deutlich gemacht, wie die Algorithmen und Techniken mit den angegebenen Datensätzen oder Eingangsdaten implementiert wurden?

- Gab es Komplikationen mit den ursprünglichen Metriken oder Techniken, die vor dem Erwerb einer Lösung geändert werden mussten?

- Gab es einen Teil des Kodierungsprozesses (z.B. das Schreiben komplizierter Funktionen), der dokumentiert werden sollte?\_

*### Refinement*In this section, you will need to discuss the process of improvement you made upon the algorithms and techniques you used in your implementation. For example, adjusting parameters for certain models to acquire improved solutions would fall under the refinement category. Your initial and final solutions should be reported, as well as any significant intermediate results as necessary. Questions to ask yourself when writing this section:  
- \_Has an initial solution been found and clearly reported?\_  
- \_Is the process of improvement clearly documented, such as what techniques were used?\_  
- \_Are intermediate and final solutions clearly reported as the process is improved?\_  
  
In diesem Abschnitt müssen Sie den Verbesserungsprozess diskutieren, den Sie mit den Algorithmen und Techniken, die Sie in Ihrer Implementierung verwendet haben, durchgeführt haben. So würden beispielsweise die Anpassung der Parameter für bestimmte Modelle zur Erlangung verbesserter Lösungen unter die Kategorie Verfeinerung fallen. Ihre Anfangs- und Endlösungen sollten gemeldet werden, ebenso wie alle wesentlichen Zwischenergebnisse, falls erforderlich. Fragen, die Sie sich beim Schreiben dieses Abschnitts stellen sollten:

- Wurde eine erste Lösung gefunden und eindeutig gemeldet?

- Ist der Verbesserungsprozess klar dokumentiert, z.B. welche Techniken verwendet wurden?\_

- Werden Zwischen- und Endlösungen eindeutig gemeldet, wenn der Prozess verbessert wird?\_

*## IV. Results*\_(approx. 2-3 pages)\_  
  
*### Model Evaluation and Validation*In this section, the final model and any supporting qualities should be evaluated in detail. It should be clear how the final model was derived and why this model was chosen. In addition, some type of analysis should be used to validate the robustness of this model and its solution, such as manipulating the input data or environment to see how the model’s solution is affected (this is called sensitivity analysis). Questions to ask yourself when writing this section:  
- \_Is the final model reasonable and aligning with solution expectations? Are the final parameters of the model appropriate?\_  
- \_Has the final model been tested with various inputs to evaluate whether the model generalizes well to unseen data?\_  
- \_Is the model robust enough for the problem? Do small perturbations (changes) in training data or the input space greatly affect the results?\_  
- \_Can results found from the model be trusted?\_

In diesem Abschnitt sollten das endgültige Modell und alle unterstützenden Eigenschaften im Detail bewertet werden. Es sollte klar sein, wie das endgültige Modell abgeleitet wurde und warum dieses Modell gewählt wurde. Darüber hinaus sollte eine Art Analyse verwendet werden, um die Robustheit dieses Modells und seiner Lösung zu validieren, wie z.B. die Manipulation der Eingangsdaten oder der Umgebung, um zu sehen, wie sich die Lösung des Modells auswirkt (dies wird als Empfindlichkeitsanalyse bezeichnet). Fragen, die Sie sich beim Schreiben dieses Abschnitts stellen sollten:

- Ist das endgültige Modell sinnvoll und auf die Lösungserwartungen abgestimmt? Sind die endgültigen Parameter des Modells angemessen?

- Wurde das endgültige Modell mit verschiedenen Eingaben getestet, um zu beurteilen, ob das Modell gut zu unsichtbaren Daten passt?\_

- Ist das Modell robust genug für das Problem? Beeinträchtigen kleine Störungen (Änderungen) der Trainingsdaten oder des Eingabebereichs die Ergebnisse erheblich?

- Können die vom Modell gefundenen Ergebnisse vertrauenswürdig sein?

*### Justification*In this section, your model’s final solution and its results should be compared to the benchmark you established earlier in the project using some type of statistical analysis. You should also justify whether these results and the solution are significant enough to have solved the problem posed in the project. Questions to ask yourself when writing this section:  
- \_Are the final results found stronger than the benchmark result reported earlier?\_  
- \_Have you thoroughly analyzed and discussed the final solution?\_  
- \_Is the final solution significant enough to have solved the problem?\_  
  
In diesem Abschnitt sollten die endgültige Lösung Ihres Modells und seine Ergebnisse mit dem Benchmark verglichen werden, den Sie zuvor im Projekt mit einer Art statistischer Analyse festgelegt haben. Sie sollten auch begründen, ob diese Ergebnisse und die Lösung signifikant genug sind, um das im Projekt auftretende Problem gelöst zu haben. Fragen, die Sie sich beim Schreiben dieses Abschnitts stellen sollten:

- Sind die Endergebnisse besser als das zuvor berichtete Benchmark-Ergebnis?

- Haben Sie die endgültige Lösung gründlich analysiert und diskutiert?

- Ist die endgültige Lösung signifikant genug, um das Problem gelöst zu haben?

*## V. Conclusion*\_(approx. 1-2 pages)\_  
  
*### Free-Form Visualization*In this section, you will need to provide some form of visualization that emphasizes an important quality about the project. It is much more free-form, but should reasonably support a significant result or characteristic about the problem that you want to discuss. Questions to ask yourself when writing this section:  
- \_Have you visualized a relevant or important quality about the problem, dataset, input data, or results?\_  
- \_Is the visualization thoroughly analyzed and discussed?\_  
- \_If a plot is provided, are the axes, title, and datum clearly defined?\_

In diesem Abschnitt müssen Sie eine Form der Visualisierung bereitstellen, die eine wichtige Qualität des Projekts hervorhebt. Es ist viel freier, sollte aber ein signifikantes Ergebnis oder Merkmal des Problems, das Sie diskutieren möchten, angemessen unterstützen. Fragen, die Sie sich beim Schreiben dieses Abschnitts stellen sollten:

- Haben Sie eine relevante oder wichtige Qualität über das Problem, den Datensatz, die Eingabedaten oder die Ergebnisse visualisiert?

- Ist die Visualisierung gründlich analysiert und diskutiert?\_

- Wenn ein Plot erstellt wird, sind die Achsen, der Titel und der Bezugspunkt klar definiert?\_

*### Reflection*In this section, you will summarize the entire end-to-end problem solution and discuss one or two particular aspects of the project you found interesting or difficult. You are expected to reflect on the project as a whole to show that you have a firm understanding of the entire process employed in your work. Questions to ask yourself when writing this section:  
- \_Have you thoroughly summarized the entire process you used for this project?\_  
- \_Were there any interesting aspects of the project?\_  
- \_Were there any difficult aspects of the project?\_  
- \_Does the final model and solution fit your expectations for the problem, and should it be used in a general setting to solve these types of problems?\_

In diesem Abschnitt fassen Sie die gesamte End-to-End-Problemlösung zusammen und besprechen ein oder zwei bestimmte Aspekte des Projekts, die Sie für interessant oder schwierig befunden haben. Von Ihnen wird erwartet, dass Sie über das Projekt als Ganzes nachdenken, um zu zeigen, dass Sie ein gutes Verständnis für den gesamten Prozess Ihrer Arbeit haben. Fragen, die Sie sich beim Schreiben dieses Abschnitts stellen sollten:

- Haben Sie den gesamten Prozess, den Sie für dieses Projekt verwendet haben, gründlich zusammengefasst?

- Gab es irgendwelche interessanten Aspekte des Projekts?\_

- Gab es irgendwelche schwierigen Aspekte des Projekts?\_

- Entspricht das endgültige Modell und die endgültige Lösung Ihren Erwartungen an das Problem, und sollte es in einem allgemeinen Rahmen zur Lösung dieser Art von Problemen verwendet werden?

*### Improvement*In this section, you will need to provide discussion as to how one aspect of the implementation you designed could be improved. As an example, consider ways your implementation can be made more general, and what would need to be modified. You do not need to make this improvement, but the potential solutions resulting from these changes are considered and compared/contrasted to your current solution. Questions to ask yourself when writing this section:  
- \_Are there further improvements that could be made on the algorithms or techniques you used in this project?\_  
- \_Were there algorithms or techniques you researched that you did not know how to implement, but would consider using if you knew how?\_  
- \_If you used your final solution as the new benchmark, do you think an even better solution exists?\_

In diesem Abschnitt müssen Sie eine Diskussion darüber führen, wie ein Aspekt der von Ihnen entworfenen Implementierung verbessert werden kann. Betrachten Sie zum Beispiel, wie Ihre Implementierung allgemeiner gestaltet werden kann und was geändert werden müsste. Sie müssen diese Verbesserung nicht vornehmen, aber die möglichen Lösungen, die sich aus diesen Änderungen ergeben, werden berücksichtigt und mit Ihrer aktuellen Lösung verglichen/kontrastiert. Fragen, die Sie sich beim Schreiben dieses Abschnitts stellen sollten:

- Gibt es weitere Verbesserungen, die an den Algorithmen oder Techniken vorgenommen werden könnten, die Sie in diesem Projekt verwendet haben?\_

- Gab es Algorithmen oder Techniken, die Sie erforscht haben, die Sie nicht implementieren konnten, die Sie aber in Betracht ziehen würden, wenn Sie wüssten, wie?\_

- Wenn Sie Ihre endgültige Lösung als neuen Maßstab verwendet haben, glauben Sie, dass es eine noch bessere Lösung gibt?  
-----------  
  
\*\*Before submitting, ask yourself. . .\*\*  
  
- Does the project report you’ve written follow a well-organized structure similar to that of the project template?  
- Is each section (particularly \*\*Analysis\*\* and \*\*Methodology\*\*) written in a clear, concise and specific fashion? Are there any ambiguous terms or phrases that need clarification?  
- Would the intended audience of your project be able to understand your analysis, methods, and results?  
- Have you properly proof-read your project report to assure there are minimal grammatical and spelling mistakes?  
- Are all the resources used for this project correctly cited and referenced?  
- Is the code that implements your solution easily readable and properly commented?  
- Does the code execute without error and produce results similar to those reported?

\*\*Vor dem Einreichen fragen Sie sich selbst. . .\*\*

- Folgt der von Ihnen erstellte Projektbericht einer gut organisierten Struktur, die der der Projektvorlage ähnelt?

- Ist jeder Abschnitt (insbesondere \*\*Analyse\*\* und \*\*Methodik\*\*) klar, präzise und spezifisch geschrieben? Gibt es mehrdeutige Begriffe oder Phrasen, die einer Klärung bedürfen?

- Wäre die Zielgruppe Ihres Projekts in der Lage, Ihre Analysen, Methoden und Ergebnisse zu verstehen?

- Haben Sie Ihren Projektbericht richtig Korrektur gelesen, um sicherzustellen, dass es nur minimale Grammatik- und Rechtschreibfehler gibt?

- Sind alle für dieses Projekt verwendeten Ressourcen korrekt zitiert und referenziert?

- Ist der Code, der Ihre Lösung implementiert, leicht lesbar und richtig kommentiert?

- Führt der Code fehlerfrei aus und liefert ähnliche Ergebnisse wie die gemeldeten?

AB HIER IST DAS PROPOSAL

## Domain Background

Due to their devastating consequences, earthquakes and their prediction are one of the most important problems in the geosciences. The National Earthquake Information Center registers between 12000-14000 earthquakes per year[2]. The number of fatalities varies greatly. Between 2000 and 2015, an average of 50100 people dies each year[3]. The economic losses, which amount to billions every year, have not yet been taken into account.

Even if earthquakes cannot be prevented, a more precise prediction of when, where and how strong an earthquake will be would be of great help. This would increase the warning time and reduce the number of fatalities.

## Problem Statement

So far there are no reliable predictions for the occurrence of earthquakes. So far it has been possible to circle the location and strength of earthquakes, but not the time. After an earthquake, data records usually show small aftershocks and certain patterns. The problem is that the sensors often measure the same thing without a dangerous earthquake following. No method is so sophisticated that earthquakes such as thunderstorms can be predicted.

To be precise, this is about predicting when an earthquake will occur. This task is a regression problem. The input is seismic data and the output is the time in seconds until the next laboratory earthquake.

## Datasets and Inputs

#### The data sets are freely available for download on the Kaggle competition website[4].

#### The data consists of a train.csv file containing a single, continuous training segment of experimental data. And test data in the test.zip file.

#### Input Data fields

* acoustic\_data - the seismic signal [int16]
* time\_to\_failure - the time (in seconds) until the next laboratory earthquake [float64]
* seg\_id - the test segment ids for which predictions should be made (one prediction per segment)
* Total rows in this train dataset: 629145480.

#### Test data set

* test.zip containing many small segments of test data.
* test dataset that is split into 2624 files , each one containing 150,000 acoustic\_data points

## Solution Statement

There are several possible options to solve the problem. It is possible to solve approaches with a "classical" ML-algorithm like Random Forrest or Linear Regression as well as with a neural net. Since the seismic signals are similar to audio signals, methods of audio signal analysis can also be used. First, an MFCC (Mel Frequency Cepstral Coefficients)[5] is used to check whether some MFCC values have a linear relationship to time\_to\_failure. These values can be used for training. The predictions determined are then executed and evaluated on the test data set.

## Benchmark Model

Since this task is a regression problem, a K-Nearst Neighbors will be used as a benchmark. The KNN is applied to the same data set to have a comparison value with the final / optimized model. The aim of benchmarking is to obtain a comparison value using a model that is as simple as possible in order to confirm that a problem can really be solved. This can also be a naive solution (guessing the mean).

As this is a Kaggle Competition, the Public Leaderboard[6] is a additional benchmark for the final solution. Due to the restriction of a maximum of 2 submissions per day, part of the training data can be used as test data to check the effectiveness of the solution compared to others.

## Evaluation Metrics

The used metric is the Mean Absolute Error[7]. The Mean Absolute Error measures the average magnitude of the errors in a set of predictions, without considering their direction. It’s the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight.

= Number of predicted values  
 = real value  
 = predicted value

For each record oft he test dataset the prediction error is calculated. Convert each error to a positive figure by taking the absolute value for each error. Finally, calculate the mean value for all recorded absolute errors.

## Project Design

I am planning the following workflow to solve the problem:

* Data preprocessing
  + Visualize a section of the data to improve the understanding of the data (using plots)
  + Apply a mel-frequency cepstral coefficients (MFCC) to identify relationships between the acoustic data and time\_to\_failure
  + Use Prinipal Component Analysis (PCA)[8] to reduce the number of trainable features
  + Splitting the data into a training and test set
* Training
  + testing different ML algorithms (regression, trees, etc.) possibly also a neural network
  + Evaluate the results to select the best algorithm
* Optimization / Finetuning
  + GridSearch to determine the best hyper parameters[9]
  + Improve performance with Cross Validation[10]

*[1]* [*https://www.kaggle.com/c/LANL-Earthquake-Prediction/*](https://www.kaggle.com/c/LANL-Earthquake-Prediction/) *[2]* [*https://www.iris.edu/hq/inclass/fact-sheet/how\_often\_do\_earthquakes\_occur*](https://www.iris.edu/hq/inclass/fact-sheet/how_often_do_earthquakes_occur) *[3]* [*https://www.statista.com/statistics/263108/global-death-toll-due-to-earthquakes-since-2000/*](https://www.statista.com/statistics/263108/global-death-toll-due-to-earthquakes-since-2000/) *[4]* [*https://www.kaggle.com/c/LANL-Earthquake-Prediction/data*](https://www.kaggle.com/c/LANL-Earthquake-Prediction/data) *[5]* [*https://en.wikipedia.org/wiki/Mel-frequency\_cepstrum*](https://en.wikipedia.org/wiki/Mel-frequency_cepstrum) *[6]* [*https://www.kaggle.com/c/LANL-Earthquake-Prediction/leaderboard*](https://www.kaggle.com/c/LANL-Earthquake-Prediction/leaderboard) *[7]* [*https://www.statisticshowto.datasciencecentral.com/absolute-error/*](https://www.statisticshowto.datasciencecentral.com/absolute-error/) *[8]* [*https://towardsdatascience.com/a-step-by-step-explanation-of-principal-component-analysis-b836fb9c97e2*](https://towardsdatascience.com/a-step-by-step-explanation-of-principal-component-analysis-b836fb9c97e2) *[9]* [*https://scikit-learn.org/stable/modules/grid\_search.html*](https://scikit-learn.org/stable/modules/grid_search.html) *[10]* [*https://towardsdatascience.com/cross-validation-in-machine-learning-72924a69872f*](https://towardsdatascience.com/cross-validation-in-machine-learning-72924a69872f)