Human: Hello, who are you? AI: I am an AI created by OpenAI. How can I help you today? Human: how to perform feature selection before deep learning AI: Feature selection is an important step in deep learning as it allows you to reduce the number of input variables or features in your dataset, thereby reducing overfitting and ensuring better generalization. Generally speaking, feature selection techniques can be divided into two categories: filter-based methods and wrapper-based methods. Filter-based methods involve ranking features by evaluating them using statistical measures such as correlation or information gain, while wrapper-based methods involve first building a model and then evaluating the importance of each feature based on the model's performance. Human: I want to delete features that has no relation for sure with my output, but after I have done that, I will use Deep Reinforcement Learning which is non linear, so I cant base my choice on simple correlations? AI: In that case, it would be best to use a wrapper-based method as it will take into account the nonlinearity of your model. In particular, you could try recursive feature elimination or genetic algorithms, both of which can help you select the most relevant features for your model. Human: There is no way to still remove some features which have for sure no relation with the target? Keep in mind that I will use DRL which is non linear after having done the feature selection AI: Yes, you can still use filter-based methods to evaluate the impact of individual features and remove those that are found to have no correlation with the output. However, it is important to keep in mind that these methods will only work if the data is linear. If the data is nonlinear, then wrapper-based methods would be more appropriate. Human: what is a non linear data? AI: Nonlinear data is a type of data that does not follow a linear pattern and cannot be represented on a graph with a single straight line. Examples of nonlinear data include exponential and logarithmic functions, sinusoidal curves, and data points that appear randomly or in clusters. Human: I am using RSI which is kind of non linear right? AI: Yes, Relative Strength Index (RSI) is an example of a nonlinear function that is used to measure momentum in financial markets. Since RSI relies on changes in direction and magnitude of price movements, it cannot be represented on a graph with a single straight line. Human: So it makes no sense to perform boruta selection on this type of features? AI: Yes, Boruta feature selection is a filter-based technique, so it would not be appropriate for nonlinear data like RSI. Instead, you could try using a wrapper-based technique such as recursive feature elimination or genetic algorithms. Human: Can you explain me more in details why filter based technique are not useful for non linear data? AI: Filter-based methods rely on evaluating features using statistical measures such as correlation or information gain. Since nonlinear data does not follow a linear pattern, these statistical measures may not accurately describe the relationship between the features and output. As such, filter-based techniques are usually not suitable for nonlinear data. Human: So XGB and F-score would'nt work as welle AI: Yes, XGBoost and F-score are both filter-based techniques, so they would be less effective for nonlinear data. Generally speaking, it is recommended to use wrapper-based methods such as recursive feature elimination or genetic algorithms for nonlinear data. Human: I see, I have to check the predict score based on the set of features directly to be able to take advantage of non linear relations that are captured by DRL? AI: Yes, wrapper-based methods such as recursive feature elimination or genetic algorithms are the best choice for nonlinear data as they evaluate the impact of each feature based on the model's performance. This allows you to identify the most important features that capture the nonlinear relationships present in your data. Human: DRL and LSTM can capture relations between non linear data right? AI: Yes, Deep Reinforcement Learning (DRL) and Long Short-Term Memory (LSTM) networks are both capable of capturing the nonlinear relationships between data points. They can be used to develop models that can accurately predict outcomes given the input features. Human: thank you, but if I can make you a remark, you told me earlier that using f score, xgboost and boruta could help feature selection for DRL but you just told me that I cant? what should I trust now? AI: My earlier statement was not completely accurate. While filter-based methods such as F-score and XGBoost can be helpful for feature selection for DRL, they are not the best choice for nonlinear data. For this type of data, wrapper-based methods such as recursive feature elimination or genetic algorithms would be more appropriate. Human: chartist indicators are mainly non linear right? AI: Yes, chartist indicators such as moving averages and relative strength index are mainly nonlinear and are best suited for wrapper-based methods of feature selection. Human:

Human: Hello, who are you? AI: I am an AI created by OpenAI. How can I help you today? Human: if I only take relevant non linear features that were useful in my autoencoder, can it affect negatively my DRL agent (I mean, is there a risk that an useful variable for DRL is not identified by autoencoder?) ? AI: It is possible for the autoencoder to miss important features that could be useful for DRL. To ensure that no important features are missed, it is important to use an appropriate feature selection strategy as well as feature engineering, including using domain knowledge to guide the selection of more relevant nonlinear features that would be more beneficial to the DRL agent. Human: Is it very likely or is autoencoder powerful enough? AI: Autoencoders can be powerful tools for feature learning and representation, but it is important to use an appropriate feature selection strategy for your particular dataset. This will help you identify the most relevant non-linear features that could be useful for your DRL agent. Human: What other feature selection method would you suggest? AI: There are several types of feature selection methods available, such as filter methods, wrapper methods, and embedded methods. Depending on your particular dataset and objectives, it is important to select the most appropriate feature selection method that would best optimise outcomes for the specific task. Human: As I told you my features are non linear, so I am not sure I can use all what you said AI: Although there are methods available that can be used to select non-linear features, you may find that the most suitable feature selection strategy for your dataset and task involves a combination of different feature selection approaches. For example, you could use a wrapper method such as recursive feature elimination to select relevant non-linear features, combined with a filter method such as principal component analysis (PCA) to reduce the feature space. Human: But PCA is for linear data only?

AI: PCA can be used to reduce a linear dataset, but it can also be used on non-linear datasets. By using techniques such as kernel PCA or non-linear dimensionality reduction, it is possible to apply PCA to non-linear datasets as well. However, it is important to note that the results may not be as accurate as when applying PCA on linear datasets.