

SIT764 Spike Report

Spike Description:

Technical spike: Explore various machine learning algorithms. To determine whether support vector machine learning algorithm offers the best classification performance on training the data.

Spike Reference

The article I have referenced is: <https://www.dezyre.com/article/top-10-machine-learning-algorithms/202>

Pros and Cons

Naïve Bayes Classifier Algorithm

Pros: performs well when the input variables are categorical.
converges faster, requiring relatively little training data than other discriminative models like logistic regression, when Naïve Bayes conditional independence assumption holds.
Cons: the classification parameter, attributes which describe the instances should be conditionally independent.

K Means Clustering Algorithm

Pros: clustering computes faster than hierarchical clustering for large number of variables.
Cons: Different initial partitions can result in different final clusters.

Support Vector Machine Learning Algorithm:

Pros: offers best classification performance and more efficiency
Cons: Long training times and difficult to understand final model.

Decision Trees

Pros: can handle for numerical and categorical inputs and is useful in data exploration.
Cons: When decision trees are made in real time the outcome may not be same as expected or planned

Recommendation:

We can use support vector algorithm or decision tree machine learning algorithm to estimate the sustainability index.