

Capstone Project Ideas

Sample Machine Learning Tasks for Capstone Project

Find a dataset, use the [New York Taxi 2013 dataset](#), or use the New York Taxi data in conjunction with other data (weather, social events, etc) and complete machine learning tasks.

Example Task 1: Binary Classification Problem.

Predict whether a tip was paid for a trip, i.e. a *tipamount that is greater than \$0 is a positive example, while a tipamount of \$0 is a negative example.*

Example Task 2: Multiclass Classification Problem.

Predict the range of tip paid for the trip. We divide the tip_amount into five bins or classes:

Class 0 : *tipamount = \$0* Class 1 : *tipamount > \$0 and tipamount <= \$5* Class 2 : *tipamount > \$5 and tipamount <= \$10* Class 3 : *tipamount > \$10 and tipamount <= \$20* Class 4 : *tipamount > \$20*

Example Task 3: Regression Task

Predict the amount of tip paid for a trip.

Example Task 4: Regression Task

Predict the travel time for a trip.

Possible Sources of Data

1. Kaggle provides a searchable list of many data sets from many sources:
<https://www.kaggle.com/datasets>
2. The University of California at Irvine (UCI) Machine Learning Repository:
<https://archive.ics.uci.edu/ml/datasets.html>
3. KD Nuggets provides a high level index to many public data sources:
<http://www.kdnuggets.com/datasets/index.html>
4. Stanford University provides another high-level index to many public data sources
<http://statweb.stanford.edu/~sabatti/data.html>
5. Bien data competitions.
<https://biendata.com/>