**Capstone Project Ideas**

1. **Cardiotocography Data Set**

* 2126 fetal cardiotocograms (CTGs) were automatically processed and the respective diagnostic features measured.
* The CTGs were also classified by three expert obstetricians and a consensus classification label assigned to each of them.
* Classification was both with respect to a morphologic pattern (A, B, C. ...) and to a fetal state (N, S, P). Therefore, the dataset can be used either for 10-class or 3-class experiments.

**Link**: [https://archive.ics.uci.edu/ml/datasets/Cardiotocography#](https://archive.ics.uci.edu/ml/datasets/Cardiotocography)

1. **FMA: A Dataset for Music Analysis Data Set**

* Audio track (encoded as mp3) of each of the 106,574 tracks. It is on average 10 million samples per track.
* Nine audio features (consisting of 518 attributes) for each of the 106,574 tracks.
* Given the metadata, multiple problems can be explored: recommendation, genre recognition, artist identification, year prediction, music annotation, unsupervised categorization.

**Link:** <https://archive.ics.uci.edu/ml/datasets/FMA%3A+A+Dataset+For+Music+Analysis>

1. **Caravan Insurance Challenge: Identify potential purchasers of caravan insurance policies**

* This data set used in the Coil 2000 Challenge contains information on customers of an insurance company.
* The data consists of 86 variables and includes product usage data and socio-demographic data derived from zip area codes.
* The data was collected to answer the following question: Can you predict who would be interested in buying a caravan insurance policy and give an explanation why?

**Link:** <https://www.kaggle.com/uciml/caravan-insurance-challenge>