**The Environmental Effect on Milk Production**

**Abstract**

The food industry is growing rapidly. Therefore, making prediction for future is a necessary task to make sure that the supply continues and reduce any possible lose or waste. Dairy is one of the sectors that is based on milk production.

The USDA wanted to make prediction of the milk production based on many factors that include the farm area, the number of operations in the state, the number of cows, the amount of calf on feed. This can be combined with seasonal changes such as the amount of rainfall, and the temperature that affect the production amount (Collier, 2017).

Due to climate change, the USDA wanted a model that can predicate the production and draw conclusions from it.

**DATA**

I scraped data from two websites, the Census of Agriculture and the weather.com. Then, continued with the process of EDA and started the regression process. Cleaning data was easy because I have implemented the try catch in my code to know precisely where the scrape might go wrong.

**Features**

Choosing the right features was managed previously; however, I have noticed multicollinearity that needed to be fixed during the process of correlation. For instance, I have removed the number of cows because its correlation was 0.99.

**Linear Regression**

I needed to use polynomial features because the score on linear regression was low 0.44. Therefore, using polynomial improved the R^2 and scored 0.90 on the train and 0.93 on the validation.

In the end of this project, I have created a model that can predict the milk production based on factors and know what is necessary to increase the process.