

**Entity Relationship Diagram**

The diagram above depicts an entity relationship diagram that encapsulates the relationship of various tables for the database. In this case, Paper\_ID from the Papers Table acts as a Primary key, similarly, Model\_ID and Variable\_ID also act as primary keys in Model and Variable tables respectively. There are junction tables named “Paper\_model table” and “Paper\_Variable” table that establishes a many-to-many relationship between various tables. In these tables ‘Paper\_ID’, ‘Model\_ID’ and ‘Variable\_ID’ all act as foreign keys.

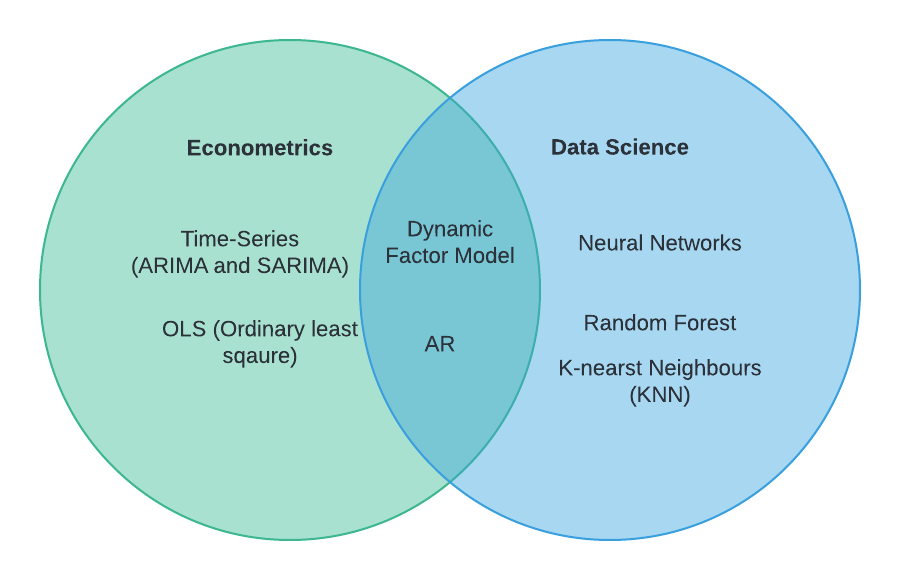
However, there have been challenges that have been faced while creating this database. One main challenge is that the variable table contains multiple variables having similar names, however, the data sources are different in many cases. Also, for some papers data sources were mentioned however, the names of the variables used were not mentioned. Considering these factors creating a complete dataset is a difficult task.

**The issue with identifying Variables**

In a lot of journal articles that use data for Neural Networks or machine learning the data sources or the variables used are missing. Complete information about either two is not available which makes creating a database for variables a difficult task.

**Relationship b/w Econometric and Data Science Modelling**

It was observed in many papers that before using machine learning models a base was set using econometric models such as different types of Autogestion model (VAR).

[](https://machinelearningmastery.com/autoregression-models-time-series-forecasting-python/#:~:text=Autoregression%20is%20a%20time%20series,range%20of%20time%20series%20problems.)

* [AR- Auto Regression](https://www.investopedia.com/terms/a/autoregressive.asp)
* [Dynamic Factor Model](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjZj82kzPiBAxUWa_EDHe-kAR0QFnoECBoQAQ&url=https%3A%2F%2Fcran.r-project.org%2Fweb%2Fpackages%2Fdfms%2Fvignettes%2Fdynamic_factor_models.pdf&usg=AOvVaw3fWxupIdG33d3_iwg80aWZ&opi=89978449)
* [OLS (Ordinary least square)](https://medium.com/analytics-vidhya/ordinary-least-square-ols-method-for-linear-regression-ef8ca10aadfc)
* [ARIMA](https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp)
* [SARIMA](https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/)
* [Neural Networks](https://www.ibm.com/topics/neural-networks)
* [K-nearest Neighbors](https://www.ibm.com/topics/knn#:~:text=Next%20steps-,K%2DNearest%20Neighbors%20Algorithm,of%20an%20individual%20data%20point.)
* [Random Forest](https://www.ibm.com/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems.)