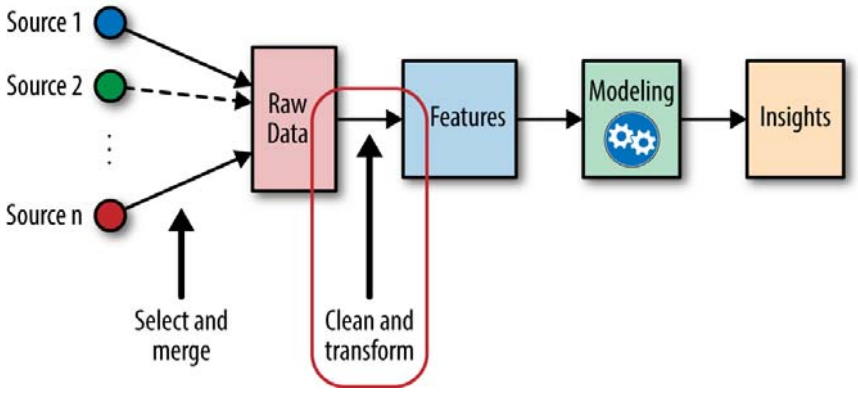
Current topic for a Master's Thesis

Feature Engineering for Ionospheric Space Weather Prediction with Machine Learning

The performance and the reliability of GNSS applications can be impacted by the Total Electron Content (TEC) of the Earth's upper atmosphere (ionosphere). The variability of TEC in the ionosphere means a manifestation of space weather. In order to minimize this effect on GNSS applications, an accurate modelling and prediction is needed.

Artificial intelligence, such as Machine Learning (ML), has shown the ability to find and learn complex patterns from historical data to solve problems. The key ingredient for a successful ML model is data, which need to be related to the problem and be able to capture the knowledge required to solve the problem. It is essential to have good quality data, large enough, relevant to the specific problem and similar to the operational data.



Feature engineering within the machine learning workflow (Zheng and Cesari, 2018)

Feature engineering is the process of transforming raw data into appropriate features better representing the underlying problem, here ionospheric space weather modelling. Feature engineering mainly has two goals, (1) preparing the proper input datasets, compatible with the requirements of the ML problem and algorithm, and (2) improving the performance of ML models by properly selecting features. Feature engineering comprises different processes that can be applied depending on the ML problem and the algorithms such as feature selection, feature extraction, feature scaling, feature transformation, etc. When features are well selected, even simpler and less optimal ML models can learn structural patterns in the data and provide realistic results. In that case it is possible to use less complex models which run faster, are easier to understand and easier to maintain.

Within this master thesis, the ML process shall be implemented with special attention to feature engineering, which will be performed on the input datasets (observations of solar activity, magnetic field, etc.) to represent the problem of ionospheric space weather modelling. At the end, data will be trained and evaluated by estimating model accuracy on unseen data.

Main tasks:

- Preparation of data (data cleaning and preprocessing) and transformation into features.
- Feature importance and feature extraction: estimation of usefulness of a feature applying correlation coefficients and methods such as Random Forest, Gradient Boosted Machines etc.; feature extraction algorithms and dimensionality reduction if needed.
- ML model training and evaluation on unseen data using the selected features.

References:

Camporeale, E., Wing, S., Johnson, J.: Machine Learning Techniques for Space Weather, 1st edition, ISBN: 9780128117880, 2018

Zheng, A., Casari, A.: Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists, O'Reilly Media, Sebastopol, CA, ISBN: 9781491953242, 2018

Institute: Deutsches Geodätisches Forschungsinstitut der TUM (DGFI-TUM)
Supervisors: Prof. Dr. Michael Schmidt / M.Sc. Randa Natras
Contact: mg.schmidt@tum.de; randa.natras@tum.de; phone 089/23031-1123