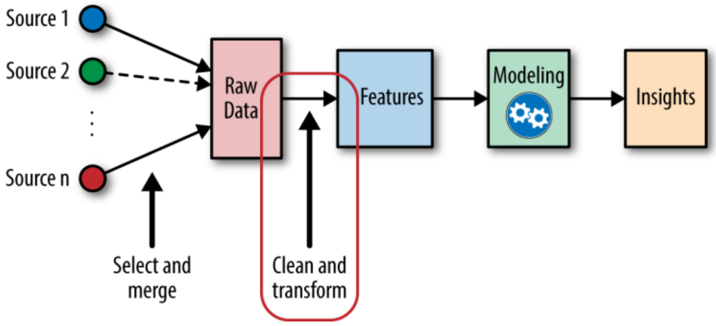
Current topic for a Master's Thesis

Feature Extraction for Ionospheric Space Weather Forecasting 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 (SW). In order to minimize SW impacts on GNSS applications, an accurate modelling and forecasting is needed.

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 able to capture the knowledge required to solve the problem. It is essential to have good quality data, large enough, relevant to the problem and similar to the operational data.



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

Feature extraction is the process of transforming raw data into appropriate features better representing the underlying problem, here SW forecasting. Its two main goals are to (1) prepare the proper input datasets, compatible with the requirements of the ML problem and learning algorithm, and to (2) improve the performance of ML models by properly selecting important features. Feature extraction comprises different processes that can be applied depending on the ML problem and learning algorithms such as feature selection, feature scaling, feature transformation. When features are well selected and prepared, even simpler ML models can learn structural patterns and provide realistic results. In that case it is possible to use less complex models that run faster, are easier to understand and to maintain.

The ML process shall be implemented with special attention to feature extraction, to be performed on the input dataset (solar / magnetic field observations) to represent the problem of SW. At the end, data will be trained and evaluated by estimating model accuracy on unseen data.

Main tasks:

- Data preprocessing (handling outliers and missing values) and transformation into features.
- Feature extraction: deriving new features from existing ones using feature extraction techniques, estimating feature usefulness applying cross-corelation and feature importance methods of Gradient Boosted Trees, Random Forest, Shapley additive explanations, etc.
- ML model training and evaluation on unseen data using the selected features.

References:

Natras, R., Soja, B., Schmidt, M.: Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting. *Remote Sensing,* 10.3390/rs14153547, 2022

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