Data Preprocessing for Supervised Learning

The purpose of this paper is to highlight best practices in data preprocessing to optimize the generalization performance of supervised learning algorithms. The paper was structured to include the preprocessing steps in order from first to last necessary during a machine learning project and goes into further detail of each step. The paper provides an overview of a preprocessing pipeline which leads to higher quality models with optimized generalization. As it goes through each step of the pipeline, it presents the 'best' algorithms at every section to achieve the best performance on a dataset.

**Part B**

This paper was insightful in terms of fundamentals and that it is certainly a piece to keep saved to refer to. Without a high-standard data preprocessing workflow, the modeling process is useless similar to saying "Garbage in, garbage out'. A model will only be as good as the data you feed it, so it is important to focus on the steps mentioned in this article rather than on the modeling process.

Exploring Millions of Footprints in Location Sharing Services

The purpose of the second paper was to explore Location Services data for insights into the spatial-temporal and social behaviors of humans. The study was first started through the gathering of data through various sources and preprocessing to filter noise. Once the data was cleaned, a spatial-temporal analysis was done to reveal the average daily and weekly check in patterns. Three major variables were used to deduce human mobility patterns: user displacement, radius of gyration and returning probability. Through these variables, the researchers were able to find that LSS users follow simple pattern and social status is an important factor in mobility.

**Part B**

I found it interesting that both the distribution of displacements and radius of gyration follow a power law scaling which could be used for other insights. Another great insight I found was that people with different mobility patterns significantly talk about different things as if perspective changes the more you travel. The way the results and methodology were presented was clear and invited reproducibility. This article opened up a new perspective in how to quantify societies as spatial-temporal networks and how we may be able to use graph neural networks to predict new hubs for location placing (predicting new node creation temporally).