**Classification of Glass Types using Neural Networks**

# Introduction

Glass is a highly versatile material used across various industries due to its unique physical and chemical properties. Its applications span from architectural windows and automotive windshields to optical lenses, containers, and tableware. Given the diversity of uses, accurately classifying glass types based on their composition is a critical task in industries like manufacturing, recycling, and quality assurance.

The classification of glass is not only important for ensuring that the correct type of glass is used in specific applications but also plays a vital role in sustainability. For instance, during recycling processes, mixing different types of glass can lead to product defects or inefficiencies in production. Furthermore, precise classification ensures adherence to safety standards, such as in automotive or building windows, where certain types of glass are required to meet stringent regulations.

In this project, we utilize a publicly available dataset that contains nine numerical features related to the refractive index and the oxide composition of glass samples. These features include the proportions of Sodium (Na), Magnesium (Mg), Aluminum (Al), Silicon (Si), and other elements critical to glass composition. The target variable categorizes the glass samples into seven distinct types, such as building windows (float or non-float processed), containers, tableware, and headlamps.

Machine learning methods, especially neural networks, are well-suited for solving classification problems due to their ability to model complex, non-linear relationships in data. Dense Neural Networks (DNNs) are particularly effective for tabular datasets like this one, where the relationships between features and the target variable are intricate.

The primary objective of this project is to develop a DNN model that accurately predicts the type of glass based on its composition. This involves exploratory data analysis (EDA) to uncover patterns in the data, preprocessing to ensure the model is fed with optimized inputs, and hyperparameter tuning to maximize the model's performance. By the end of this project, the goal is to deliver a robust model that can classify glass types with high accuracy while providing insights into the strengths and challenges of using neural networks for such tasks.

# Problem Description

The classification of glass types poses a unique challenge due to the inherent complexity of the material's composition and its applications across diverse industries. Glass is made from a mixture of raw materials such as sand, soda ash, and limestone, combined with specific oxides to achieve desired properties. These compositions determine the glass's refractive index, durability, transparency, and thermal properties, which are critical for its intended use. For instance, the glass used in building windows requires a different manufacturing process and composition than glass used for headlamps or containers.

In the context of the provided dataset, this problem is framed as a multi-class classification task where the objective is to predict one of seven distinct glass types based on its compositional features. These types include categories like float-processed building windows, non-float-processed building windows, containers, tableware, and headlamps. Each class represents a specific industrial application, with precise compositional requirements.

Several challenges make this problem particularly complex:

1. Overlapping Feature Distributions: Some glass types have very similar oxide compositions, leading to significant overlap in feature distributions. This can make it difficult for traditional models to differentiate between classes, particularly for classes like building windows and tableware.
2. Class Imbalance: The dataset exhibits an imbalance in class representation, with certain types of glass (e.g., headlamps) being underrepresented. Such imbalances can lead to biased models that favor the majority classes while underperforming on minority classes.
3. Multicollinearity: Features like Calcium (Ca) and Sodium (Na) exhibit high correlation, which introduces redundancy in the data. Multicollinearity can impact the model's ability to correctly attribute importance to individual features, potentially leading to suboptimal classification performance.
4. Feature Scaling: The numerical features in the dataset, such as oxide proportions and refractive index, vary widely in scale. Models sensitive to feature magnitudes, such as neural networks, require normalization to ensure that no single feature disproportionately influences the predictions.
5. Data Representation: The dataset is composed of purely numerical features, with no categorical or contextual data. While this simplifies preprocessing, it also means the model must rely solely on compositional differences, without access to auxiliary information such as production methods or physical characteristics.