## Industry 4.0

Implementation of AI for the Industry 4.0 in a fiber optic sensor production process

## **Data Exploration**

#### **Data Exploration**

I explored the characteristics and patterns of the data, including its:

- size
- shape
- distribution
- missing values
- data types

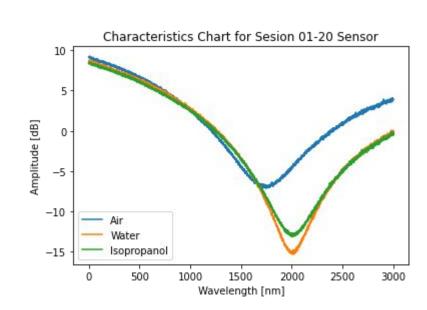
Based on my exploration, I have identified potential issues or challenges in the data that required preprocessing steps.

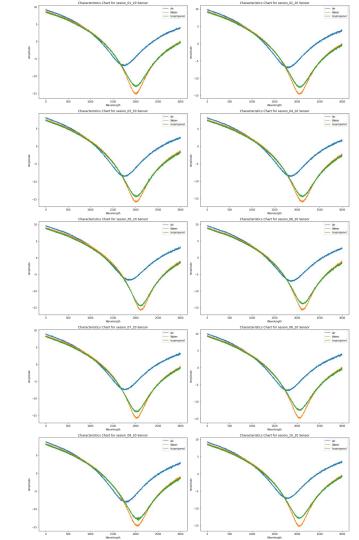
### Data Visualization

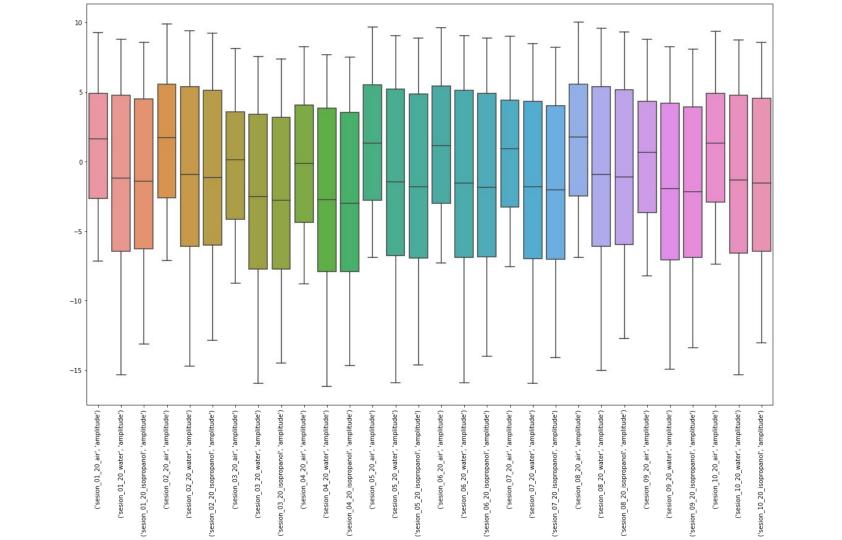
#### **Data Visualization**

Data visualization is a powerful tool to explore and understand patterns, relationships, and trends in data.

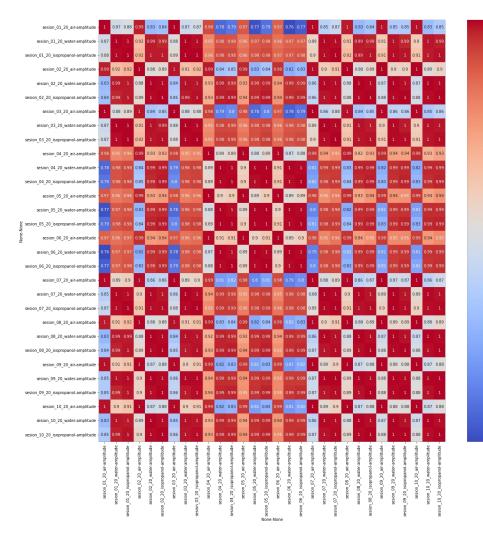
By visualizing data I gain insights into the distributions of variables, correlations between variables, and the overall structure of the data.







#### Correlation



- 0.95

- 0.90

- 0.85

- 0.80

Correlation heatmap for air amplitude and water amplitude	-1.00		Correlation heatmap for air	amplitude and isopropanol amplitude
sesion_01_20_air-amplitude -1.00 0.99 1.00 0.98 0.97 0.97 1.00 1.00 1.00 1.00 0.87 0.83 0.87 0.78 0.77 0.76 0.85 0.83 0.85 0.83		sesion_01_20_air-amplitude - 1.00 0.9	99 1.00 0.98 0.97 0.97 1.00 1.00 1.00	0 1.00 0.88 0.84 0.87 0.79 0.79 0.77 0.87 0.84 0.85 0.85
sesion_02_20_air-amplitude -0.99 1.00 1.00 0.99 0.99 0.99 1.00 1.00		sesion_02_20_air-amplitude - 0.99 1.0	00 1.00 0.99 0.99 0.99 1.00 1.00 1.00	0 1.00 0.92 0.89 0.92 0.85 0.84 0.83 0.91 0.89 0.90 0.90
sesion_03_20_air-amplitude -100 100 100 098 098 097 100 100 100 100 088 084 088 079 078 078 086 084 086 085		sesion_03_20_air-amplitude - 1.00 1.0	00 100 0.98 0.98 0.97 1.00 1.00 1.00	0 1.00 0.89 0.85 0.88 0.80 0.80 0.79 0.88 0.85 0.86 0.86
sesion_04_20_air-amplitude - 0.98 0.99 0.98 1.00 1.00 1.00 0.99 0.99 0.99 0.95 0.93 0.95 0.89 0.88 0.87 0.94 0.92 0.94 0.93		sesion_04_20_air-amplitude - 0.98 0.9	99 0.98 1.00 1.00 1.00 0.99 0.99 0.99	9 0.99 0.96 0.93 0.95 0.89 0.89 0.88 0.95 0.93 0.94 0.93
sesion_05_20_air-amplitude - 0.97 0.99 0.98 100 100 100 0.98 0.99 0.99 0.99 0.95 0.93 0.96 0.90 0.89 0.89 0.95 0.93 0.94 0.94	- 0.95	sesion_05_20_air-amplitude - 0.97 0.9	99 0.98 100 100 100 0.98 0.99 0.99	9 0.99 0.96 0.94 0.96 0.90 0.90 0.89 0.96 0.94 0.95 0.94
sesion_06_20_air-amplitude - 0.97 0.99 0.97 1.00 1.00 1.00 0.98 0.99 0.99 0.99 0.96 0.94 0.96 0.91 0.90 0.89 0.95 0.94 0.95 0.94		sesion_06_20_air-amplitude - 0.97 0.9	99 0.97 1.00 1.00 1.00 0.98 0.99 0.99	9 0.99 0.97 0.94 0.96 0.91 0.91 0.90 0.96 0.95 0.95 0.95
sesion_07_20_air-amplitude -1.00 1.00 1.00 0.99 0.98 0.98 1.00 1.00 1.00 0.89 0.86 0.89 0.81 0.80 0.79 0.88 0.86 0.87 0.86		sesion_07_20_air-amplitude - 1.00 1.0	00 1.00 0.99 0.98 0.98 1.00 1.00 1.00	0 1.00 0.90 0.86 0.90 0.82 0.81 0.80 0.89 0.87 0.87 0.87
sesion_08_20_air-amplitude -1.00 1.00 1.00 0.99 0.99 0.99 1.00 1.00		sesion_08_20_air-amplitude - 1.00 1.0	00 1.00 0.99 0.99 0.99 1.00 1.00 1.00	0 1.00 0.92 0.88 0.91 0.84 0.84 0.83 0.91 0.89 0.89 0.89
sesion 09_20_air-amplitude -100 100 100 0.99 0.99 0.99 1.00 1.00 1.0	- 0.90	sesion_09_20_air-amplitude - 100 10	00 1.00 0.99 0.99 0.99 1.00 1.00 1.00	0 1.00 0.91 0.88 0.91 0.83 0.83 0.82 0.90 0.88 0.88 0.88
sesion 10 20 air-amplitude -100 100 100 0.99 0.99 0.99 100 100 100 100 0.90 0.87 0.90 0.82 0.81 0.81 0.89 0.87 0.88 0.87	None	sesion_10_20_air-amplitude - 1.00 1.0	00 1.00 0.99 0.99 0.99 1.00 1.00 1.00	0 1.00 0.91 0.88 0.91 0.83 0.83 0.82 0.90 0.88 0.88 0.88
sesion 01 20 water-amplitude -0.87 0.92 0.88 0.95 0.96 0.96 0.89 0.91 0.91 0.90 1.00 0.99 1.00 0.98 0.97 0.97 1.00 0.99 1.00 1.00	g sesion_0	1_20_isopropanol-amplitude - 0.88 0.9	02 0.89 0.96 0.96 0.97 0.90 0.92 0.93	1 0.91 100 100 100 0.98 0.98 0.98 1.00 1.00 1.00 1.00
sesion 02 20 water-amplitude -0.83 0.88 0.84 0.93 0.93 0.94 0.86 0.88 0.87 0.87 0.99 1.00 1.00 0.99 0.99 0.99 1.00 1.00	sesion_0	2_20_isopropanol-amplitude - 0.84 0.8	39 0.85 0.93 0.94 0.94 0.86 0.88 0.8	8 0.88 100 100 100 0.99 0.99 0.99 100 100 100 100
sesion 03 20 water-amplitude -0.87 0.91 0.88 0.95 0.96 0.96 0.89 0.91 0.90 0.90 1.00 1.00 0.98 0.98 0.98 1.00 1.00 1.00 1.00	_	3_20_isopropanol-amplitude - 0.87 0.9	02 0.88 0.95 0.96 0.96 0.90 0.91 0.91	1 0.91 100 100 100 0.98 0.98 0.98 1.00 1.00 1.00 1.00
sesion 04 20 water-amplitude -0.78 0.84 0.79 0.89 0.90 0.91 0.81 0.83 0.82 0.82 0.98 0.99 0.98 1.00 1.00 1.00 0.99 0.99 0.99	- 0.85 sesion_0-	4_20_isopropanol-amplitude - 0.79 0.8	0.80 0.89 0.90 0.91 0.82 0.84 0.8	3 0.83 0.98 0.99 0.98 1.00 1.00 1.00 0.99 0.99 0.99 0.99
sesion 05 20 water-amplitude -0.77 0.83 0.78 0.88 0.89 0.90 0.80 0.82 0.81 0.81 0.97 0.99 0.98 1.00 1.00 1.00 0.98 0.99 0.99 0.99	sesion_0	5_20_isopropanol-amplitude - 0.79 0.8	0.80 0.89 0.90 0.91 0.81 0.84 0.8	3 0.83 0.98 0.99 0.98 1.00 1.00 1.00 0.99 0.99 0.99 0.99
sesion 06 20 water-amplitude -0.76 0.82 0.78 0.87 0.89 0.89 0.79 0.82 0.81 0.81 0.97 0.99 0.98 1.00 1.00 1.00 1.00 0.98 0.99 0.99 0.99	sesion_0	5_20_isopropanol-amplitude - 0.77 0.8	0.88 0.89 0.90 0.80 0.83 0.8	2 0.82 0.98 0.99 0.98 1.00 1.00 1.00 0.98 0.99 0.99 0.99
sesion 07 20 water-amplitude -0.85 0.90 0.86 0.94 0.95 0.95 0.88 0.90 0.89 0.89 1.00 1.00 1.00 0.99 0.98 0.98 1.00 1.00 1.00	sesion_0	7_20_isopropanol-amplitude - 0.87 0.9	01 0.88 0.95 0.96 0.96 0.89 0.91 0.9	0 0.90 100 100 100 0.99 0.99 0.98 100 100 100 100
sesion 08 20 water-amplitude -0.83 0.88 0.84 0.92 0.93 0.94 0.86 0.88 0.87 0.87 0.99 1.00 1.00 0.99 0.99 1.00 1.00 1.00	- 0.80 sesion_0	3_20_isopropanol-amplitude - 0.84 0.8	39 0.85 0.93 0.94 0.95 0.87 0.89 0.8	8 0.88 100 100 100 0.99 0.99 0.99 100 100 100 100
	sesion_0	9_20_isopropanol-amplitude - 0.85 0.9	0 0.86 0.94 0.95 0.95 0.87 0.89 0.8	8 0.88 100 100 100 0.99 0.99 0.99 100 100 100 100
sesion_09_20_water-amplitude -0.85 0.90 0.86 0.94 0.94 0.95 0.87 0.89 0.88 0.88 1.00 1.00 1.00 0.99 0.99 0.99 1.00 1.00	sesion_1	0_20_isopropanol-amplitude - 0.85 0.9	0 0.86 0.93 0.94 0.95 0.87 0.89 0.8	8 0.88 100 100 100 0.99 0.99 0.99 100 100 100 100
sesion_10_20_water-amplitude - 0.83 0.89 0.85 0.93 0.94 0.94 0.86 0.88 0.87 0.87 100 100 100 0.99 0.99 0.99 1.00 1.00 1.		iftude	mplitude mpl	amplitude amplitude amplitude amplitude amplitude amplitude amplitude amplitude
amplitude		ir-ampl	ampl ampl ampl ampl ampl ampl ampl ampl	air-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol-ampliol
		.20_air	on_03_20_air. on_04_20_air. on_05_20_air. on_05_20_air. on_07_20_air. on_07_20_air. on_08_20_air.	sesion_10_20_air 01_20_isopropanol 02_20_isopropanol 02_20_isopropanol 04_20_isopropanol 04_20_isopropanol 04_20_isopropanol 04_20_isopropanol 04_20_isopropanol 04_20_isopropanol 04_20_isopropanol 04_20_isopropanol
sion 01_20_air. sion 01_20_air. sion 03_20_air. sion 03_20_air. sion 03_20_air. sion 05_20_air. sion 05_20_air. sion 05_20_air. sion 05_20_air. sion 05_20_air. sion 05_20_air. n_01_20_air. n_02_20_air. n_02_20_air		ion_01_	ion_03_20_ion_03_20_ion_04_20_ion_05_20_ion_05_20_ion_06_20_ion_07_20_iion_08_20_iion_08_20_iion_09_20_iion_09_20_iion_09_20_iion_09_20_iion_09_20_iion_09_20_iion_09_20_iion_09_20_iion_00_20_20_20_iion_00_20_20_iion_00_20_20_iion_00_20_20_iion_00_20_20_iion_00_20_20_iion_00_20_20_20_iion_00_20_20_20_iion_00_20_20_20_20_20_20_20_20_20_20_20_20_	92510n, 10 01.20 [sopp. 02.20 [sopp. 03.20 [sopp. 04.20 [sopp. 05.20 [sopp. 06.20 [sopp. 07.20 [sopp. 09.20 [
sesion_01_20_air sesion_01_20_air sesion_02_20_air sesion_03_20_air sesion_06_20_air sesion_06_20_air sesion_06_20_air sesion_08_20_air no_01_20_water on_02_20_water on_04_20_water on_04_20_water on_05_20_water on_06_20_water		ši ši		sesion_10_20_air. n_01_20_isopropanol. n_02_20_isopropanol. n_03_20_isopropanol. n_04_20_isopropanol. n_05_20_isopropanol. n_05_20_isopropanol. n_05_20_isopropanol. n_05_20_isopropanol. n_05_20_isopropanol. n_08_20_isopropanol. n_08_20_isopropanol. n_08_20_isopropanol. n_10_20_isopropanol. n_10_20_isopropanol.
84 84 84 84 84 84 84 84 85 <u>55 55 55 55 55 55 55 55 55 55 55 55 5</u>				
Notie-Notie				MOIIC-MOIIC

- 0.90

- 0.85

#### **Data Visualization**

A characteristic chart of amplitude and wavelength between air, water, and isopropanol can provide insights into the behavior of these substances when subjected to a certain wavelength of electromagnetic radiation, such as light.

In general, the amplitude of the electromagnetic radiation can be thought of as the intensity of the signal, while the wavelength corresponds to the distance between consecutive peaks or troughs of the wave.

Based on the chart, we can observe that air generally has the lowest amplitude and a relatively shorter wavelength compared to water and isopropanol. This suggests that air may absorb less electromagnetic radiation at a given wavelength compared to water and isopropanol.

On the other hand, both water and isopropanol exhibit higher amplitudes and longer wavelengths compared to air, indicating that these substances may absorb more electromagnetic radiation at a given wavelength. The exact amplitude and wavelength values can depend on the specific wavelength range being considered and the concentration of the substances.

Overall, this type of chart can provide insights into the relative absorbance properties of different substances and can be useful in fields such as spectroscopy and materials science.

# Data Preprocessing and Feature Engineering

#### Data Preprocessing and Feature Engineering

The data was scaled to ensure that all the features were on the same scale. This was done using standardization technique.

```
from sklearn.preprocessing import StandardScaler
# create a standard scaler object
standard scaler = StandardScaler()
# fit and transform the data using the standard
scaler
df scaled standard =
standard scaler.fit transform(df)
```

#### Data Preprocessing and Feature Engineering

Furthermore, feature engineering techniques were used to extract relevant features from the raw data. These techniques include dimensionality reduction using principal component analysis (PCA)

```
    df_scaled_standard.shape
        (3001, 60)

[41] X_pca.shape
        (3001, 2)
```

```
from sklearn.decomposition import PCA

# create a PCA object with the desired
number of components
pca = PCA(n_components=2)

# fit the PCA object to your data
pca.fit(df_scaled_standard)

# transform your data to the new PCA space
X_pca = pca.transform(df_scaled_standard)
```

## Building Machine Learning Models

## Building Machine Learning Models

e.g. Linear Regression

```
# Prepare the data for machine learning
X = df.filter(regex=' air')
y = df.filter(regex='(water|isopropanol)')
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
scaler = StandardScaler()
X train = scaler.fit transform(X train)
X test = scaler.transform(X test)
pca = PCA(n components=2)
X train pca = pca.fit transform(X train)
X test pca = pca.transform(X test)
# Feature selection with RFE
estimator = LinearRegression()
selector = RFE(estimator, n features to select=5)
selector.fit(X train pca, y train)
X train selected = selector.transform(X train pca)
X test selected = selector.transform(X test pca)
# Train a linear regression model
model = LinearRegression()
model.fit(X train selected, y train)
# Make predictions on the test set
y pred = model.predict(X test selected)
# Evaluate the model
mse = mean_squared_error(y test, y pred)
r2 = r2 score(y test, y pred)
print(f"Mean Squared Error: {mse}")
print(f"R2 Score: {r2}")
```

Mean Squared Error: 2.671537845060998 R2 Score: 0.9428791226589753

#### **Building Machine Learning Models**

During the process of building machine learning models, several types of models were used, including linear regression, decision tree, random forest, and neural networks using the TensorFlow library. These models were used to analyze and make predictions based on the data set being used. Linear regression was used to predict a continuous variable, while decision trees and random forests were used for classification and regression problems. Neural networks were used for more complex problems and were implemented using TensorFlow. Overall, a variety of machine learning models were utilized to find the best fit for the specific problem and data set.

## **Preliminary Results**

#### **Preliminary Results**

The evaluation of the machine learning models used in predicting the characteristics of sensors in water and isopropanol revealed that the models were able to predict the characteristics with an accuracy ranging from 93% to 99%, depending on the specific model used. For model evaluation, commonly used metrics such as Mean Squared Error (MSE), R-squared (R2), and Mean Absolute Error (MSAE) were used. Also loss and score were also used as evaluation metrics.

Loss is a metric used in machine learning models to measure how well the model is able to approximate the relationship between the input variables and the output variable during training. In TensorFlow, loss is commonly used as the objective function that the optimizer tries to minimize during training.

Score is a metric used to evaluate the performance of a machine learning model on a validation or test set. The specific score metric used depends on the type of problem being solved (classification, regression, etc.) and the evaluation criteria. For example, for classification problems, commonly used score metrics include accuracy, precision, recall, and F1 score, while for regression problems, R2, MSE, and MSAE are commonly used as score metrics.

if you are interested in seeing more about this project for industry 4.0, please feel free to check out my project!

https://github.com/M1-KO/Industry-4.0