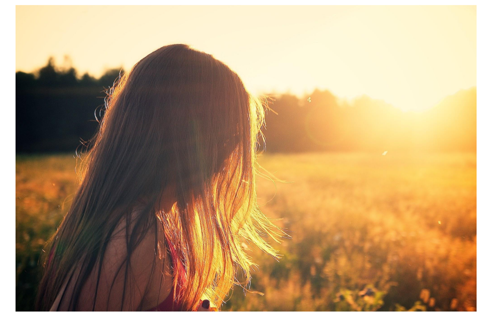
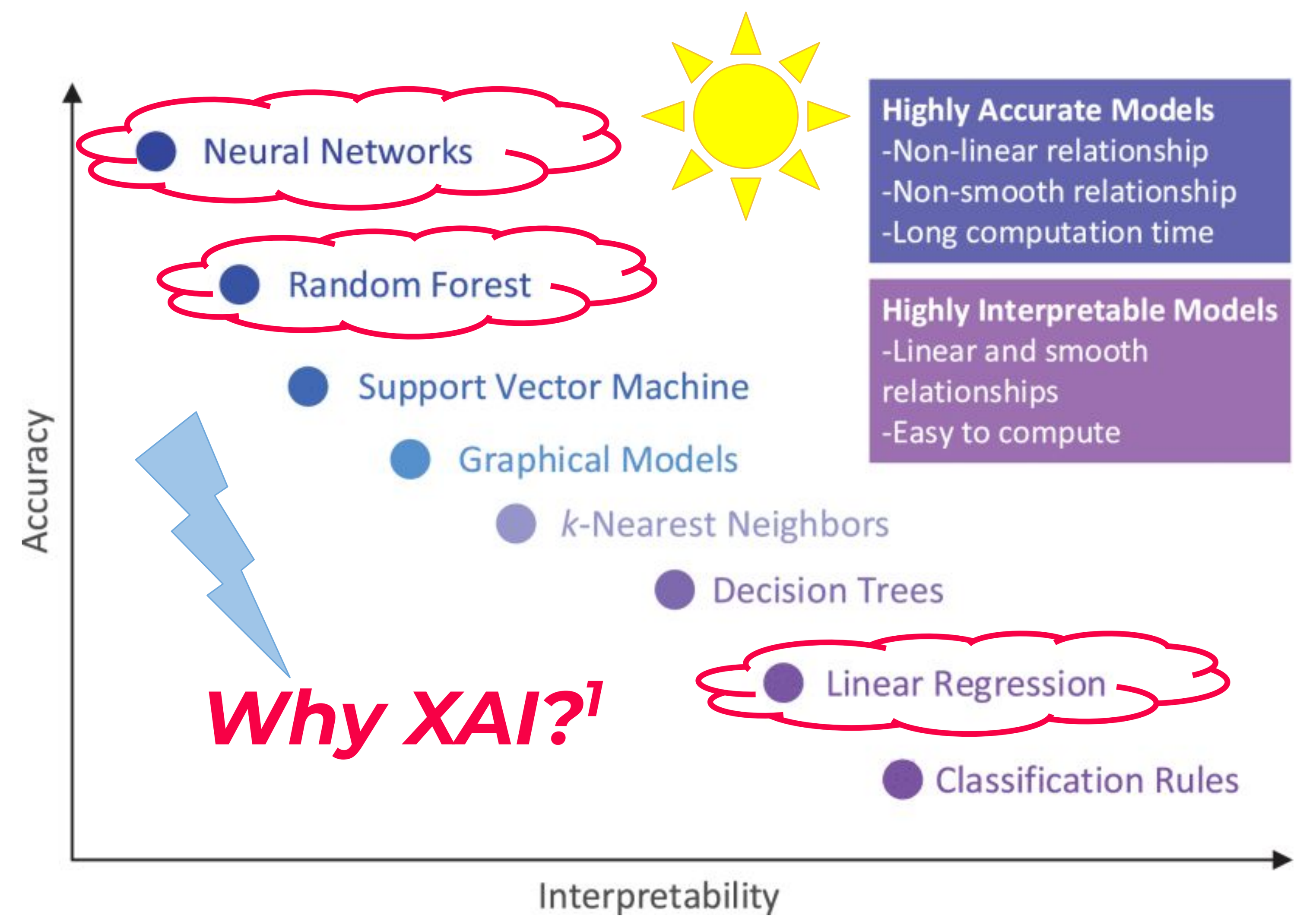
XAI: How To Look For Sunshine?

Using Explainable Artificial Intelligence (XAI) To Shed Light On The Black Box



Source: Unsplash

Naida Nožić, Dominik Homer, and Christof Haar



Problem

Compare three machine learning models (Random Forest, CNN Linear Regression) in predicting the city sunshine hours and explain the learning patterns and results.

Model decisions are **"black boxes"** and are hard to interpret.

As a solution we utilize **Explainable AI (XAI)**, which helps us understand and trust the output of AI models.

Data Understanding

→ The dataset consists of daily weather data of 15 cities.
→ Sunshine hours are the target variable for prediction.
→ 3654 rows x 165 columns (numerical features and target)
→ Preprocessing and cleaning conducted

- Data type conversion,
- Handling missing/invalid values,
- Normalization

Features
Temp mean
Temp max
Temp min
Cloud cover
Global radiation
Humidity
Pressure
Precipitation
Wind gust
Wind speed

Our Start: Linear Regression

Standard Multiple Linear Regression:
Understand how ***sunshine duration*** is influenced by other weather factors such as ***global radiation, cloud cover, and similar variables*** by fitting a linear equation to the data
→ Checked **mean squared error (MSE) and R^2**
→ Variance Inflation Factors (VIF) > 10 → **High multicollinearity**

Lasso:

+ Adds penalty to reduce coefficient values
+ sets coefficient of strongly multicollinear features to zero

− High VIF values: Tends to randomly select one feature

Ridge:

+ Penalizes large coefficient values more efficiently than Lasso

− No zero-value coefficients = no automatic feature selection

We Can Do Better: Convolutional Neural Networks (CNN)

Deep learning model that uses filters (kernels) sliding over input data (usually images) to automatically learn and detect patterns.

80-20 train-test split, manual hyperparameter tuning:

+ Better prediction power, lower multicollinearity issues

− Harder to understand

The Alternative: Random Forest (RF)

Combines multiple models to build many decision trees on random subsets of data and features, then combines their votes to improve accuracy and reduce overfitting.

80-20 train-test split, GridSearchCV-hyperparameter tuning, 5-fold cross validation (CV):

+ Better prediction power, lower multicollinearity issues
Less prone to overfitting than a single decision tree
Can capture complex non-linear relationships

− Computationally expensive for complex datasets
Harder to interpret than a single decision tree

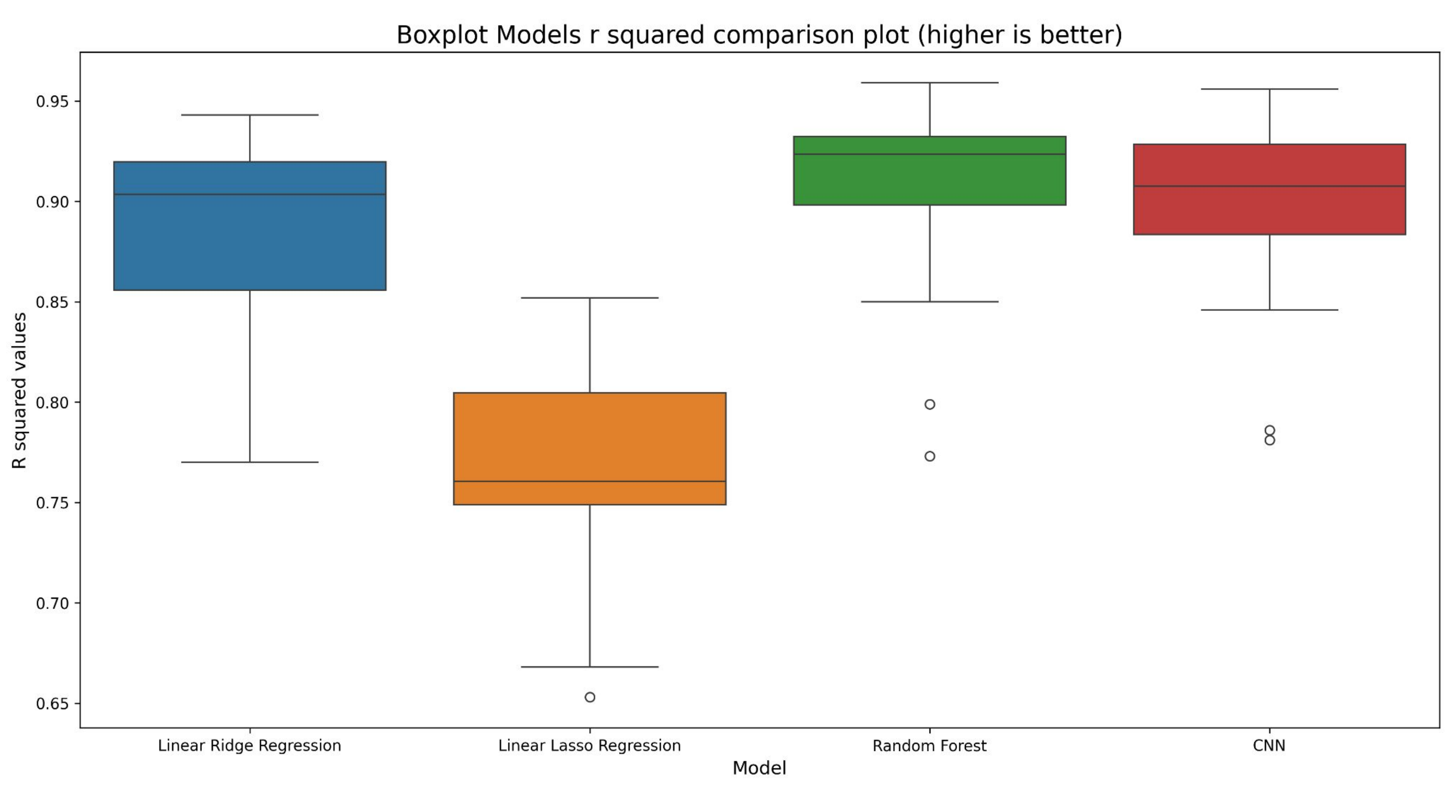
XAI: SHapley Additive exPlanations (SHAP)

Uses a game-theoretic approach to explain **how much each feature contributes to a model's prediction** (importance value for each feature) for more transparent and interpretable AI decisions.

XAI: Counterfactual Explanations

Show the **smallest changes needed in the input data that would lead a model to produce a different outcome**, helping us understand how and why decisions could be altered.

Evaluation



In terms of predicting sunshine duration based on the given data, we used **mean squared error (MSE) and R^2** values as metrics to evaluate the performance of each model. As shown in the picture above, Random Forest provided the best results in terms of the **R^2** values, while **MSE** provided very similar results.

Sensitivity Analysis

In order to test the sensitivity of the models, we changed the most relevant features ,'global_radiation', and 'humidity' by +5% and -5% and analyzed how the score metrics change by that. We did that for every model, except the Lasso model, as it's values where already much worse than the others. By that we identified which model would be most suitable for which use-case:

Ridge Regression: most robust, but worst scoring metrics overall

CNN: best balanced between robustness and capturing complex relationships

Random Forest: best at capturing complex relationships with high sensitivity, but least robust

Sources:

[1] Source: M. E. Morocho-Cayamcela, H. Lee and W. Lim, "Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions," in IEEE Access, vol. 7, pp. 137184-137206, 2019, doi: 10.1109/ACCESS.2019.2942390. Caption, sun, lightning, and clouds added.
[2] xxx
[3] xxx
[4] xxx