

Glossary

For your reference, here are all the new terms we introduced in this lesson:

- **One-hot encoding** - transforming a column of categorical data into many columns of binary values associated with the categories.
- **Deep learning embedding** - a vector of numerical values representing some input value. Similar input values will have similar embeddings. Used in text, images, and videos to represent highly dimensional data into small dimensions.
- **Null values** - values that are missing in a dataset. Represented as `None` or `NaN`.
- **R²** - measures the proportion of variance.
- **Root Mean Squared Error (RMSE)** - Measures standard deviation of prediction errors.
- **Residuals** - Differences between true target values and predicted values.
- **Accuracy** - total correctly predicted by the total number of values.
- **Precision** - the number of true positives by the number of true and false positives.
- **Recall** - the number of true positives by the number of true positives and false negatives.
- **F1** - harmonic mean of the precision and recall. Used with unbalanced datasets.
- **Hyperparameters** - model's configuration parameters which modify how the model functions.
- **Grid Search** - searches hyperparameters in a deterministic fashion across all options available for a model.
- **Randomized search** - randomly selects hyperparameters across all options available for a model.
- **Bayesian search** - uses statistical methods to optimally select hyperparameters across available options for a model.