INTRODUCTION TO Decision Tree

DECODING DATA SCIENCE

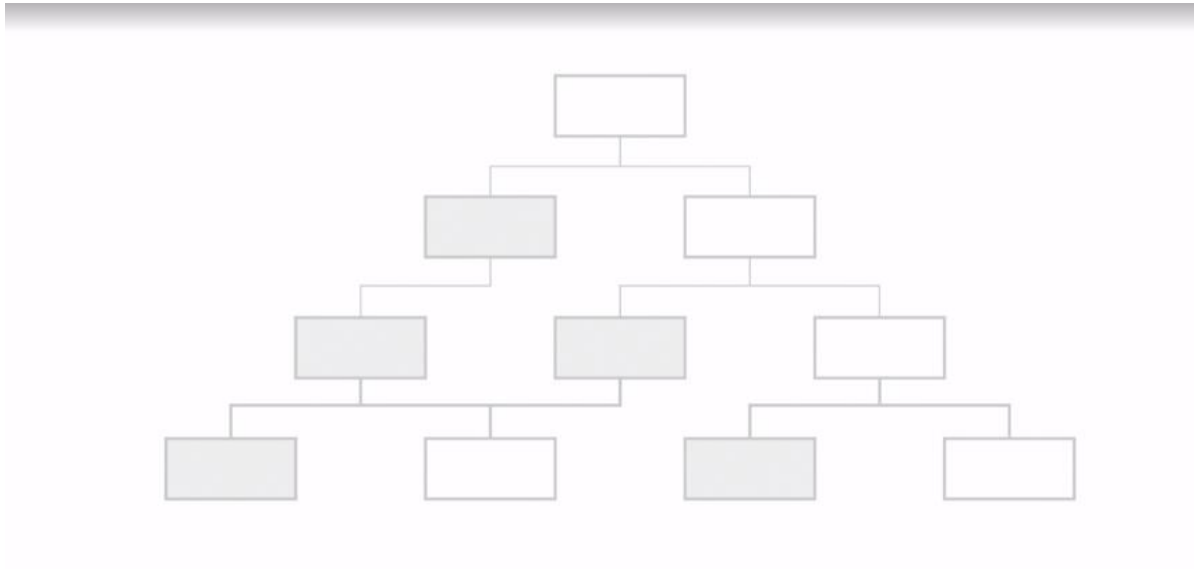
Decision Trees

A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

DECODING DATA SCIENCE

What is Decision Trees ?

A decision tree is a flowchart-like tree structure where an internal node represents a "test" on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes).



DECODING DATA SCIENCE

Importance of decision tree:

Decision trees are a popular method for various classification and regression tasks. For example, in medical diagnosis, decision trees have been used to classify diseases based on symptoms. In credit scoring, decision trees are used to predict the probability of default.

Decision tree learning is a predictive modelling approach that can be used for both classification and regression tasks. The goal is to create a model that predicts the value of a target variable based on several input variables.

Decision Trees

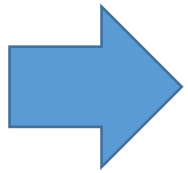
We were creating a decision tree to predict whether or not someone is likely to go to the beach.

Predictors

Sky

Weekend

Wind Speed



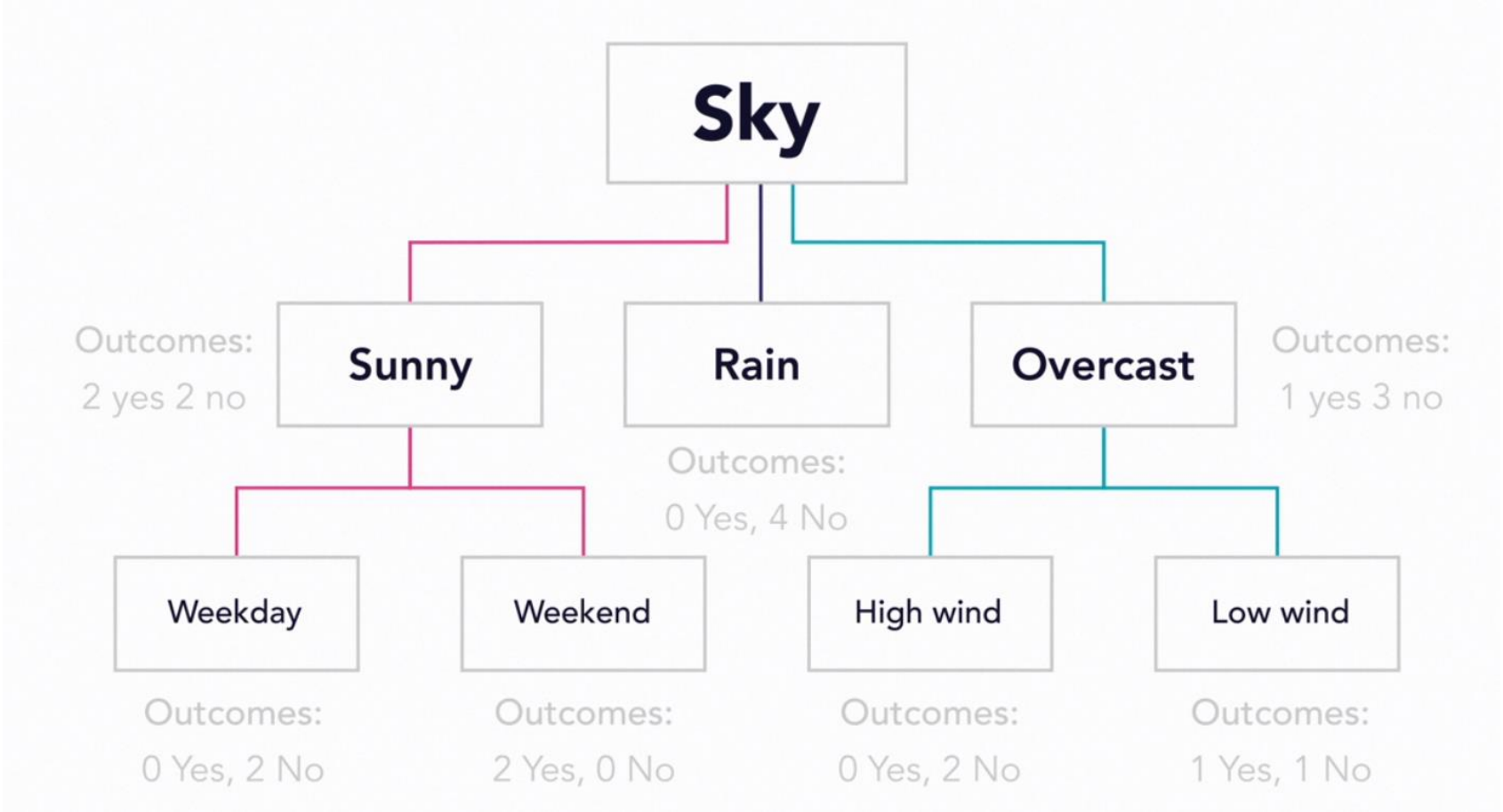
Outcome

Arshad goes to the beach?

Data

Predictors			Outcome
Sky	Weekend	Wind	Yash goes to the beach
Sunny	Weekday	Low	No
Sunny	Weekday	High	No
Overcast	Weekday	Low	Yes
Rain	Weekday	Low	No
Rain	Weekend	Low	No
Rain	Weekend	High	No
Overcast	Weekend	High	No
Sunny	Weekend	Low	Yes
Sunny	Weekend	Low	Yes
Rain	Weekend	Low	No
Overcast	Weekend	Low	No
Overcast	Weekday	High	No

Decision Tree Outcome



Advantages of decision trees:

1. Decision trees are easy to interpret and explain.

2. They can handle both numeric and categorical data.

3. They are resistance to overfitting.

4. They can be used for feature selection.

5. They are non-parametric, meaning they make no assumptions about the underlying data distribution.

Disadvantages of decision trees

1. They are prone to overfitting with large amounts of data.

2. They can be unstable because small changes in the data can result in large changes in the structure of the tree.

3. They are not as accurate as some other methods, such as neural networks.

4. They can be difficult to tune, especially when there are many parameters to consider.

6 detailed modules to be a successful data scientist

1. Excel for Data Science

- Formulas and Functions
- Visualization
- Dashboarding
- Data Analysis

2. Business Intelligence

- Tableau
- Power BI

3. Deep Dive into SQL

- Filtering Data
- Functions in Database
- Displaying Data from Multiple tables
- Grouping Data and Computing Aggregates
- Subqueries and Nested queries in SQL
- Windows Function

4. Statistics and Mathematics

- Descriptive Statistics
- Hypothesis Testing
- Fundamentals of Probability
- Linear Algebra
- ANOVA and Covariance

5. Deep Dive into Python

- Conditionals and Loops
- Operation and Operator
- Functions and Classes
- Data Wrangling – NumPy & Panda
- Visualization: Matplotlib & Seaborn

6. Decode Machine Learning

- Linear Regression
- Logistic Regression
- Decision Trees
- Clustering

Thank You!

Get in touch for Data Career

connect@decodingatascience.com