

Bird Sighting



Dheeraj Joshi
Nitish Mahipal Surana

Image Source:
Youtube -Aves BraZil
spark.apache.org

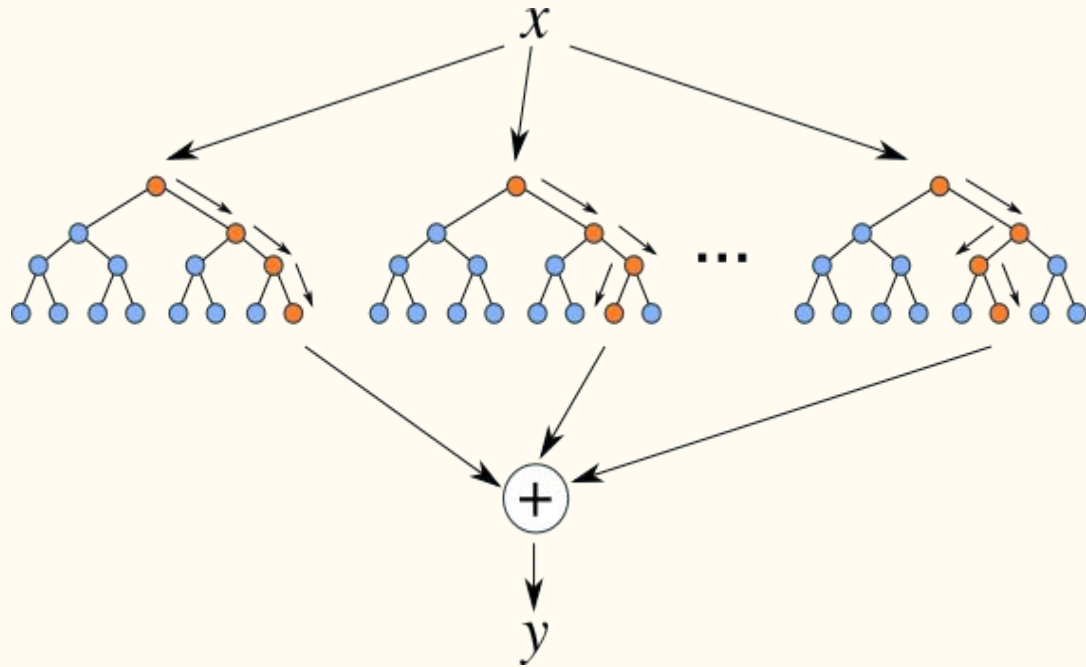
Features

- Total Features: 1658
- Core-covariates (3.8% [63]) : HOUSING_DENSITY,, HOUSING_PERCENT_VACANT, ELEV_GT, etc..
- Selected features (4.4% [71]) : DIST_FROM_WET_VEG_FRESH, DIST_FROM_FLOWING_FRESH, DIST_FROM_STANDING_FRESH, etc..
- Not selected features (95.6% [1587]) : LATITUDE, BAILEY_ECO_REGION, YEAR, COUNTRY, ELEV_GT, NUMBER_OBSERVERS, GROUP_ID, etc..

Models & Accuracy

Model	Accuracy (%)
Naive Bayes	46.72
Decision Trees	78.79
Gradient Boosted Trees	75.58
Random Forest	79.4

Random Forest



Tunable Parameters
Number of Trees
Number of Features
Depth of each Tree
Maximum Bins
Impurity
Feature Subset Strategy
Categorical Feature Mapping

Image Source: cochincodclub.wordpress.com

Random Forest Iterations

Number of Trees	Number of Features	Depth of each Tree	Feature subset strategy	Accuracy (%)	Categorical Features Used
10	11	5	auto	72.52	No
10	11	8	auto	73.31	No
10	59	8	auto	73.60	No
10	71	8	auto	74.21	No
15	71	8	auto	75.68	No
25	71	8	auto	75.6	No
20	71	8	auto	75.87	No
20	71	12	auto	77.36	No
30	71	15	sqrt	79.08	Yes
30	71	15	log2	79.09	Yes
30	71	15	all	79.40	Yes

Performance Comparison

Number of Workers	Steps	Time Taken (seconds)
5	Model Training and Validation	1114
	Prediction	158
10	Model Training and Validation	396
	Prediction	65