




CS6240 PROJECT

Biyanta Shah
Sravani Beeram



Problem



Solution

- MLLIB – Random Forests
- Why Random Forest?
- Parameters



Features used for training

- Core covariates

Included features related to climate, ecosystem and human population

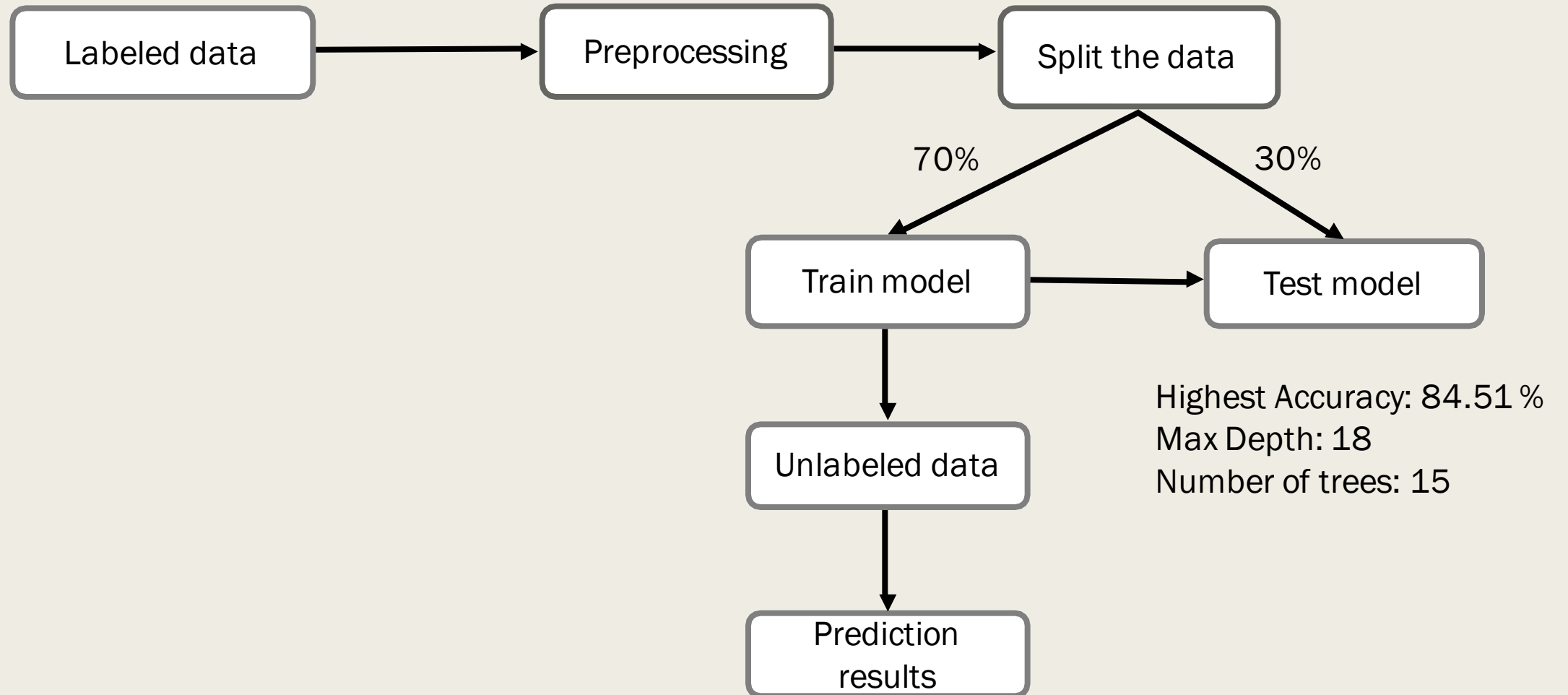
- Checklist

Included features describing the sampling events

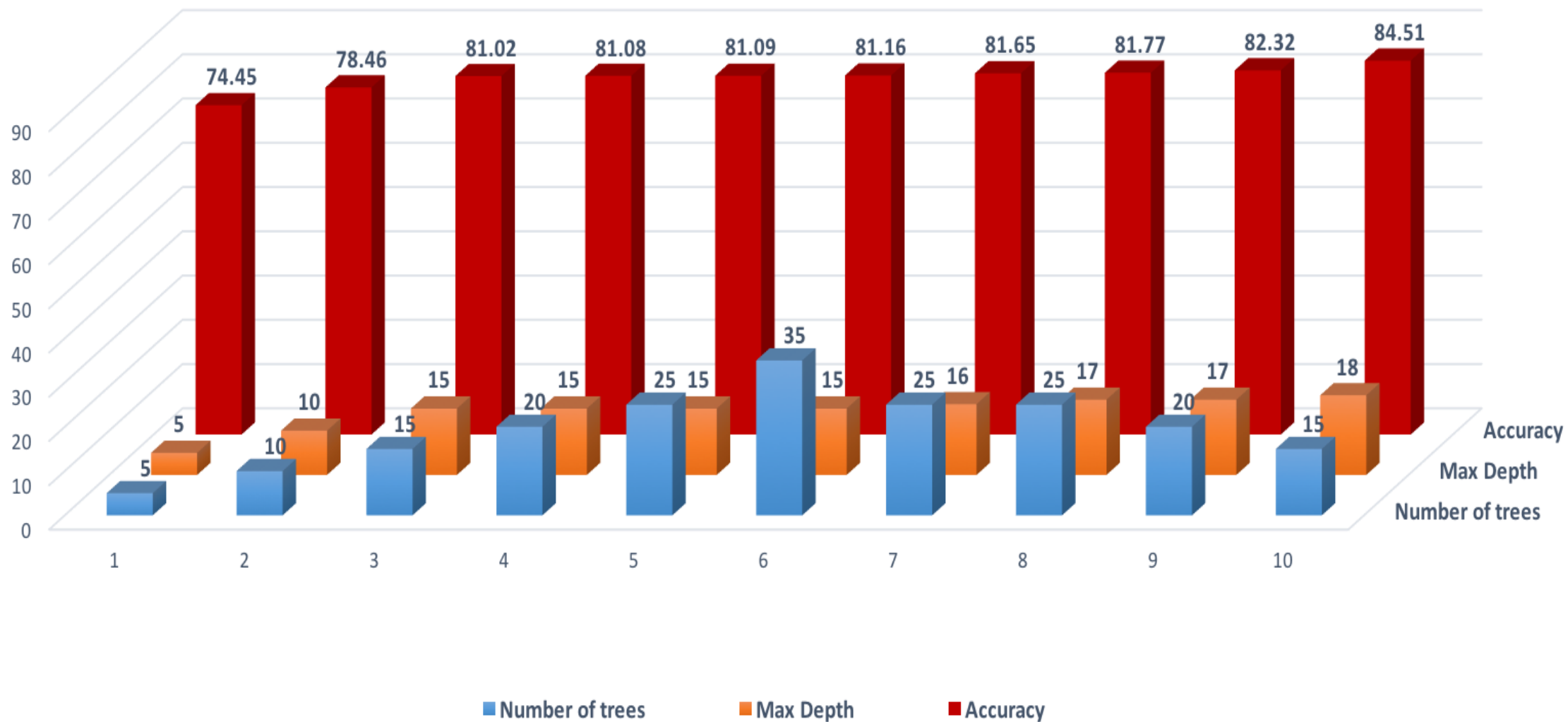
- Extended covariates

Considered features relating to habitat of the species

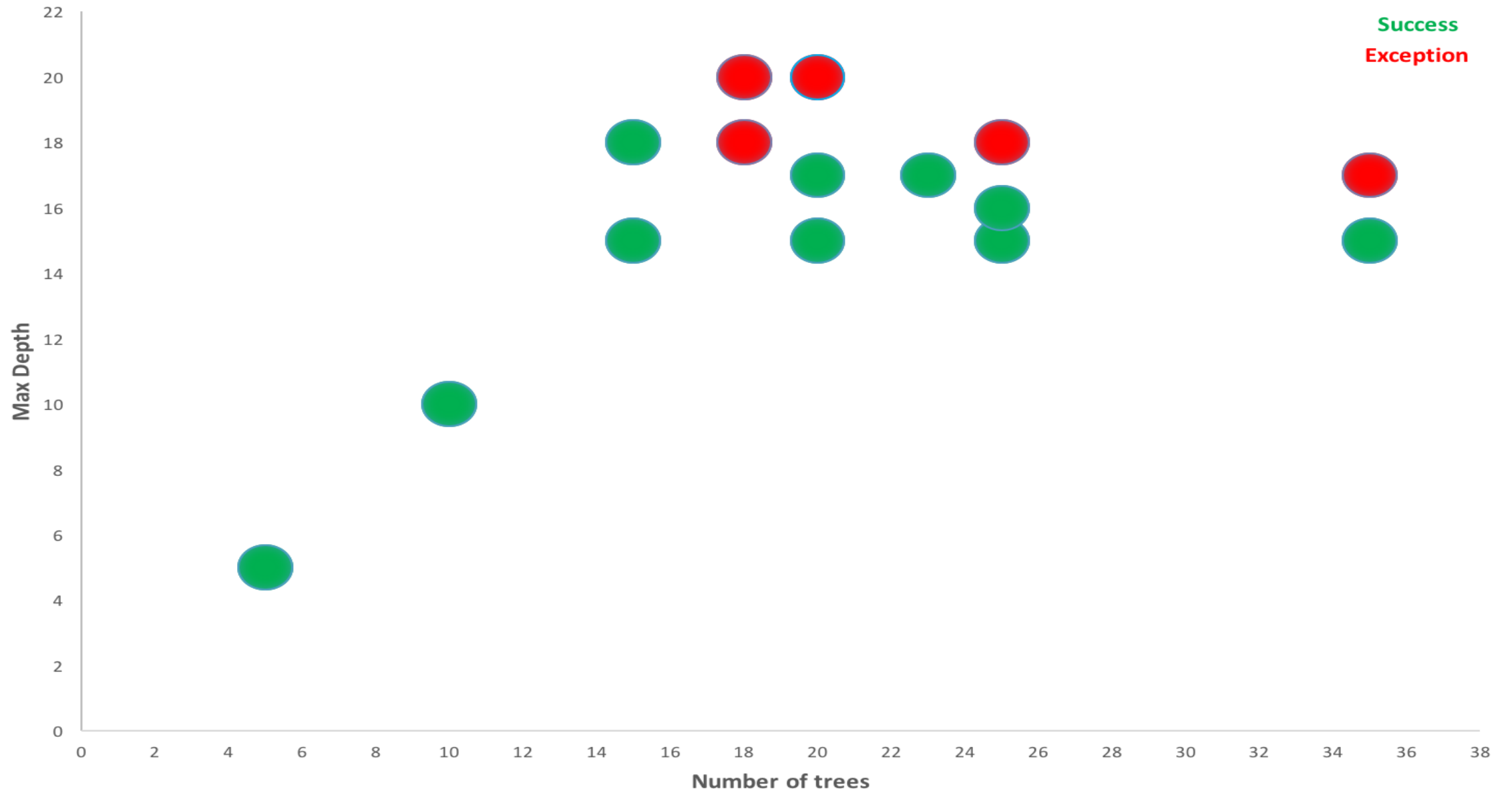
Code Design



Accuracy in relation with Number of Trees and Max Depth



RESULTS BY TUNING PARAMETERS



References

- <https://spark.apache.org/docs/latest/mllib-ensembles.html>
- <https://databricks.com/blog/2015/01/21/random-forests-and-boosting-in-mllib.html>
- <http://pages.cs.wisc.edu/~matthewb/pages/notes/pdf/ensembles/RandomForests.pdf>
- http://www.zhiyuanlin.com/uploads/5/8/8/1/58812849/improving_mllib.pdf
- <https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/>

Thank you !