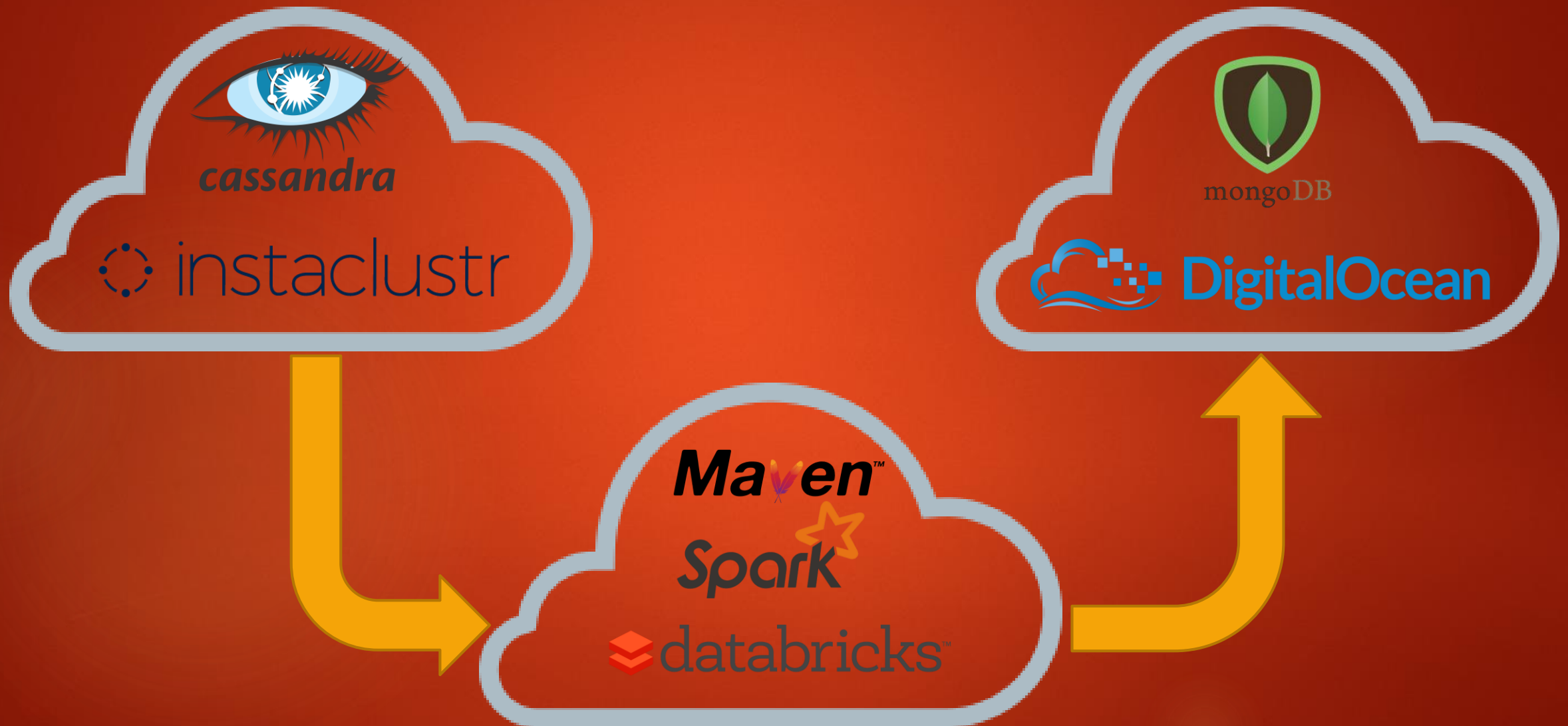


PREDICTING MEDICAL APPOINTMENT NO-SHOWS IN BRAZIL USING SPARK MLLIB



by Craig Barbisan

Target Architecture



Feature Set

Inputs to Predictive Model

Raw Features

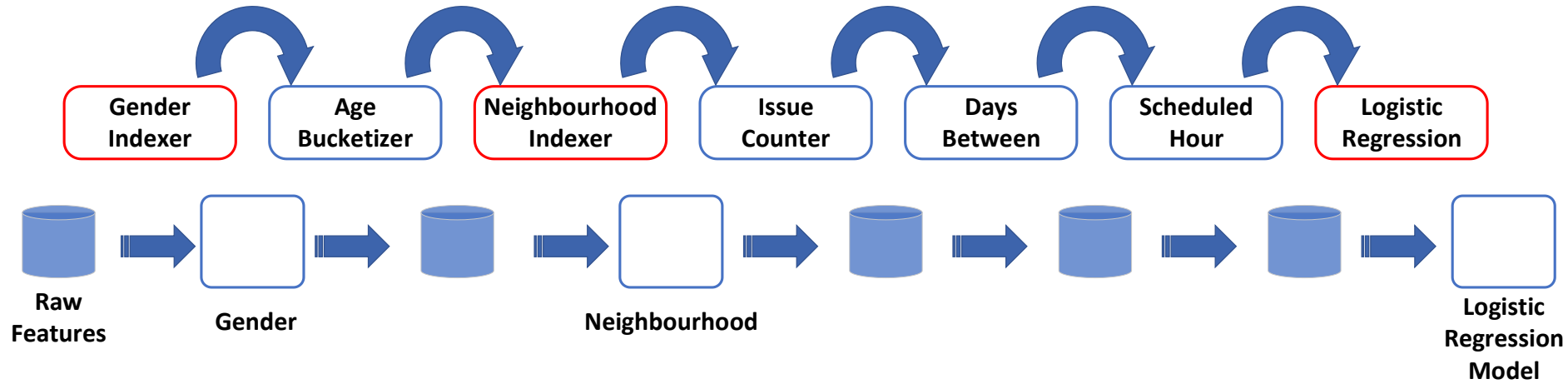
- Gender
- ScheduledDay
- AppointmentDay
- Age
- Neighbourhood
- Scholarship
- Hypertension
- Diabetes
- Alcoholism
- Handicap
- SMS_Received

Extracted Features

- GenderIndex
- AgeBucket
- NeighbourhoodIndex
- IssueCounter
- DaysBetween
- ScheduledHour

Extraneous Features

- AppointmentID
- PatientID



Training the Model

Model Evaluation

- ▶ Parameters tuned were just those specific to the logistic regression stage of the pipeline.
- ▶ Other stages did not have parameters that required tuning.
- ▶ Once the best model was selected, it was applied to the test data set.
- ▶ Comparison of the predicted value to the raw NoShow value matched 79.6% of the time.

The Best Parameters:

```
-----  
logreg_7feb07e7fa79  
res23: org.apache.spark.ml.param.ParamMap =  
{  
    logreg_7feb07e7fa79-aggregationDepth: 2,  
    logreg_7feb07e7fa79-elasticNetParam: 0.0,  
    logreg_7feb07e7fa79-family: binomial,  
    logreg_7feb07e7fa79-featuresCol: features,  
    logreg_7feb07e7fa79-fitIntercept: true,  
    logreg_7feb07e7fa79-labelCol: NoShow,  
    logreg_7feb07e7fa79-maxIter: 10,  
    logreg_7feb07e7fa79-predictionCol: prediction,  
    logreg_7feb07e7fa79-probabilityCol: probability,  
    logreg_7feb07e7fa79-rawPredictionCol: rawPrediction,  
    logreg_7feb07e7fa79-regParam: 0.01,  
    logreg_7feb07e7fa79-standardization: true,  
    logreg_7feb07e7fa79-threshold: 0.5,  
    logreg_7feb07e7fa79-tol: 1.0E-6  
}
```

Model Applications

- Resource Planning – Staff can be reduced on days when multiple no-shows are predicted.
- Waitlist Management – Additional appointments can be booked on days when no-shows are expected.
- Outreach – Patients who are expected to be no-shows could be pro-actively contacted to discuss benefits of keeping the appointment or to re-schedule.