MILESTONE 7

David Maila & Neo Ncube

BIN 381

Contents

Introduction

The implementation phase of the process is where we put what we've learned to use in the market. To deploy a model effectively, one must first comprehend the necessity. The aim of this project was to use employee demographic data to create a model that could reliably classify people with salaries above R50000. This was done in order for the sponsor to save money This was implemented so that the sponsor could also save money on his marketing strategies and target their services only on this kind of individual, knowing that this is someone who is likely to purchase.

After we have built your predictive model we deploy it in production. We give a record of features on which the model is trained and then it gives prediction as an output. The model is placed on a server where it can receive multiple requests for predictions. After you have deployed the model it is capable of taking inputs and giving output to as many different requests made to it for predictions.

Plan deployment

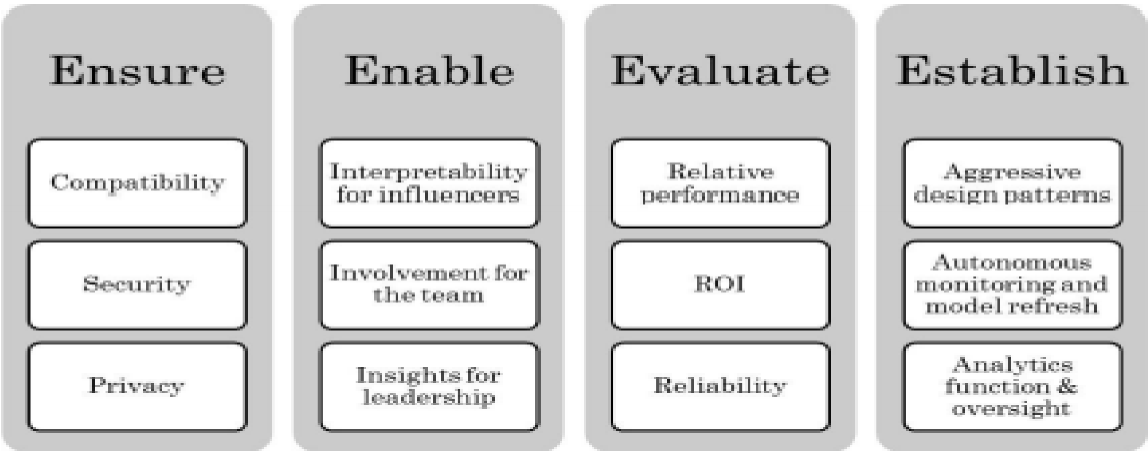

Figure 1: Sustainable deployment plan proposal (Shahapurkar, 2016)

Summarize deployable results

As the initial business objective was to select candidates for a new service offered by the sponsor of the project targeting individuals with salaries exceeding R50 00 using the decision tree as our model we discovered that people who qualify for the new service that is offered by the sponsor are the operators and videographers.

```
> confusionMatrix(prediction, test$Earns.More, positive='yes')
Confusion Matrix and Statistics

          Reference
Prediction    no   yes
       no  16539  6970
       yes 11264 22623

               Accuracy : 0.6823
                 95% CI : (0.6785, 0.6861)
    No Information Rate : 0.5156
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.361

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.7645
            Specificity : 0.5949
         Pos Pred Value : 0.6676
         Neg Pred Value : 0.7035
             Prevalence : 0.5156
         Detection Rate : 0.3942
   Detection Prevalence : 0.5904
      Balanced Accuracy : 0.6797

       'Positive' Class : yes
```

Figure 2: Confusion matrix results of model

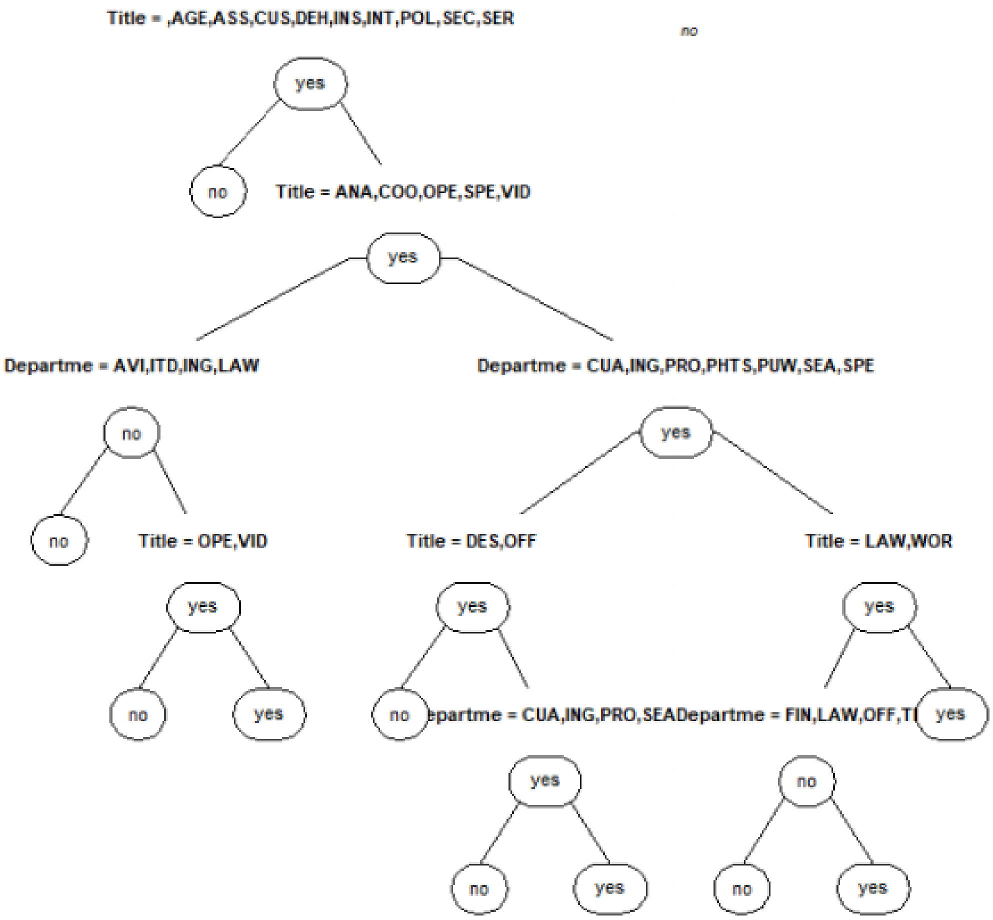

Figure 3: The tree model diagram

Develop and evaluate alternative plans for deployment

We've decided to go for a neural network. Neural networks are particularly good at generating associations that are too abstract for rule-based or clustering-based methods to handle. Neural networks find associations between a wide range of inputs and ideal outcomes? It's critical to comprehend how neural networks function, as well as how their built-in learning mechanisms are used, and how this can impact the model's implementation and upkeep.

Decide for each distinct knowledge or information result

Veteran workers are not compensated as much as new employees, which encourages new employees to stick with the firm Because people who have been with the firm for less than 2 years don't earn more 50 000 Individuals earning more than R50k are assigned to separate agencies. When R drew the tree diagram,

We may deduce from our findings that there is a strong link between employee experience and the department in which they serve, as well as whether or not they receive more than R50k.

Determine how knowledge or information will be propagated to user

We can create a website for the sponsors that gives them access to real-time reports produced by Power Bi and R. The sponsor will be able to maintain an up-to-date list of workers who earn more than R50K based on this study as employees' wages rise due to results, increased expertise, number of years served, or even a shift of offices.

Decide how the use of the result will be monitored and its benefits measured

The sponsors' use of the results will be tracked to see how easily they can get an accurate list of workers who receive more than R50K if they need it. As data analysts, our role is to provide policy makers with the information they need to make money-saving decisions.

How the model result will be deployed within the organization systems

Approaches that can be used in deployment:

Recreate

The old version is shut down; then, the new one is activated. This process is what happens when you have a simple server, and you update your web site.

Ramped

The new version is deployed and progressively delivered to the users. This solution is suggested when you are releasing a new application, and you want to tune the performance or some critical feature. It’s also a good option if you’re going to collect users’ feedback and adjust your application before the big launch.

Blue\Green

The new version is deployed alongside the other one; then, the traffic is switched to the new server. The switch happens only when the new version is online and tested, so the operation is immediate and safe.

A/B Testing

Both versions, new and old, work together. Some targeted users land on the new one so that you can get feedback

Shadow

The new version of the application receives traffic through the old platform and gives a retro compatible response to the caller.

We are going to use the blue/green approach due the reduction of downtime and risk by running two identical production environments called Blue and Green.

At any time, only one of the environments is live, with the live environment serving all production traffic. This technique can eliminate downtime due to app deployment. In addition, blue-green deployment reduces risk: if something unexpected happens with your new version on Green, you can immediately roll back to the last version by switching back to Blue.

How its use will be monitored and its benefits measured

To ensure model accuracy remains at a high percentage, the test set will continually be update to the most recent dataset yearly. This will test the model's accuracy for the most recent year's data.

➢To ensure that the attributes being used for the current model is adding to model, test will be done on the model without those attributes to check the drop in accuracy if any.

➢ if tests are done on the model and an accuracy drop of more than 30% is recorded, the training dataset will be altered to take a more recent section of the dataset and analysis will be made on this training set to try and improve the model.

➢ if attributes are to be added to the model in an attempt to improve its accuracy, those attributes will have to undergo data wrangling before being added to the dataset

Well use the Cost-benefit analysis to use measure the benefits, it compares the cost to produce the product, service, or result of the project to the benefit that the organization will receive as a result of executing the model.

Model maintenance and monitoring

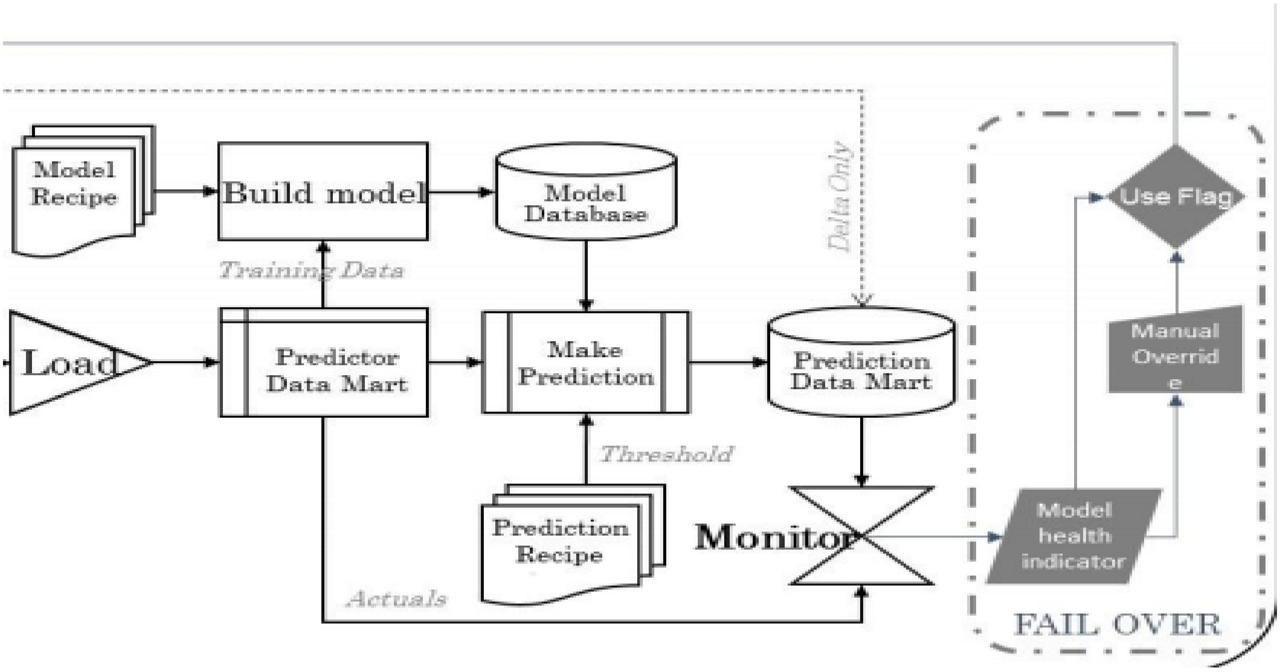

Figure 4: Model cyle and setup

Check for dynamic aspects

In real-world application, the most common aspect that change over time are (Shahapurkar, 2016):

- Predictive Accuracy: closeness of the predicted to actual responses
- Interpretability: output and operations understandable to humans
- Complexity: how compact (simple) is the learned model
- Robustness: capability of handling noise, missing values etc.
- Stability: robustness over time with the changing "world"
- Efficiency: time and memory needed for the training and test phases
- Scalability: How much the system's performance (e.g., speed) is sensitive to the size of the data set

These are the metric constantly being tweaked by the analyst or the model itself if it is self-learning. They will either improve or deteriorate.

Improvements are seen when the model determines a better percentage increase on the given aspect through continued testing and comparison with the old data and new data.

Deterioration happens when the particular aspect does not improve or gets reduce performance as a result of attempting to modify another aspect.

How accuracy will be monitored

With the advice by Demsar (J. Demšar, 2006) it is evidently shown that accuracy can be monitored in a system by employing cross validation for comparative evaluation. This process involves experimental evaluation and comparative evaluations in the learning algorithms. It can be invariably seen that these methods are applied within the following lines:

- Select an evaluation metric, the most often used one being accuracy without much thought put into the selection.

- Select a large-enough number of datasets [the number is chosen so as to be able to make a convincing case for evaluation and the datasets are generally obtained from a public data repository if not from the organisation.
- Select the best parameters for various learning algorithms, a task generally known as model tuning but unavoidably associated with evaluation.
- Use a k-fold cross-validation technique for error estimation, often stratified 10-fold cross-validation, with or without repetition.
- Apply paired t tests to all pairs of results or to the pairs deemed relevant (e.g., the ones including a possibly new algorithm of interest) to test for statistical significance in the observed performance difference.
- Average the results for an overall estimate of the algorithm's performance or, alternatively, record basic statistics such as win/loss/ties for each algorithm with respect to the others.

These step although making several assumptions with differing implications that can have several implications in real-world applications.

With the three methods available:

- Cross validation
- Bootstrapping
- Separate subset

Criteria for model abandonment, identification and recommendations

We refer to a study by Japkowiz & Shah (Japkowicz and Shah,2011) that tables the benchmarks across differing algorithms which is a table used in the cross validation to monitor the performance of the chosen model and continued relevance.

Table 1: Accuracy Measurement across Benchmark Datasets and Algorithms

Data Set	1NN	NB	BAG(REP)	SVM	C45	RIP	RF
Anneal	99.11	96.43	98.22	99.44	98.44	98.22	99.55
Audiology	75.22	73.42	76.54	81.34	77.87	76.07	79.15
Balance	79.03	72.3	82.89	82.89	76.65	81.6	80.97
Breast cancer	65.74	71.7	67.84	66.16	75.54	68.88	69.99
Contact lenses	63.33	71.76	68.33	71.67	81.67	75	71.67
Pima diabetes	70.17	74.36	74.61	77.08	73.83	75	74.88
Glass	70.5	70.36	69.63	62.21	62.21	70.95	79.87
Hepatitis	80.63	83.21	84.5	80.63	83.79	78	84.58
Hypothyroid	91.52	98.22	99.55	93.58	99.58	99.42	99.39
Tic-tac-toe	81.63	69.62	92.07	99.9	85.07	97.39	93.94
Average	77.69	78.14	81.42	82.35	81.92	82.05	83.40

We observe that only 3% of the variation (as measured by sum-of-squared variance) is contributed by the algorithm, the rest 97% is due to the datasets. This proves and show the initial premise infers that the model will only be relevant while the data set is relevant to it. A comprehensive match-making process is needed based on characteristics of the data and the demands of the domain.

Discussion on possible objective change

The initial objective as stated above was for the identification of individuals that earn more than 50000. The system is to be dynamic in order to be able to classify individuals by this criteria given their geological data. This is a target based objective and thus it factor that the business has chosen this according to some given evidence within the continual day to day processes.

It is possible to find statistical evidence upon further evaluation that the identification criteria might be second to another one. With another criteria being more compelling than the other, it will bring about reconsideration of the model. With such a change there would be a need to evaluate according to three main components of evaluation: purpose, metric and method. If the current model passes this cycle then it can be reconsidered for the changed objective by implementing small adjustments to the metrics.

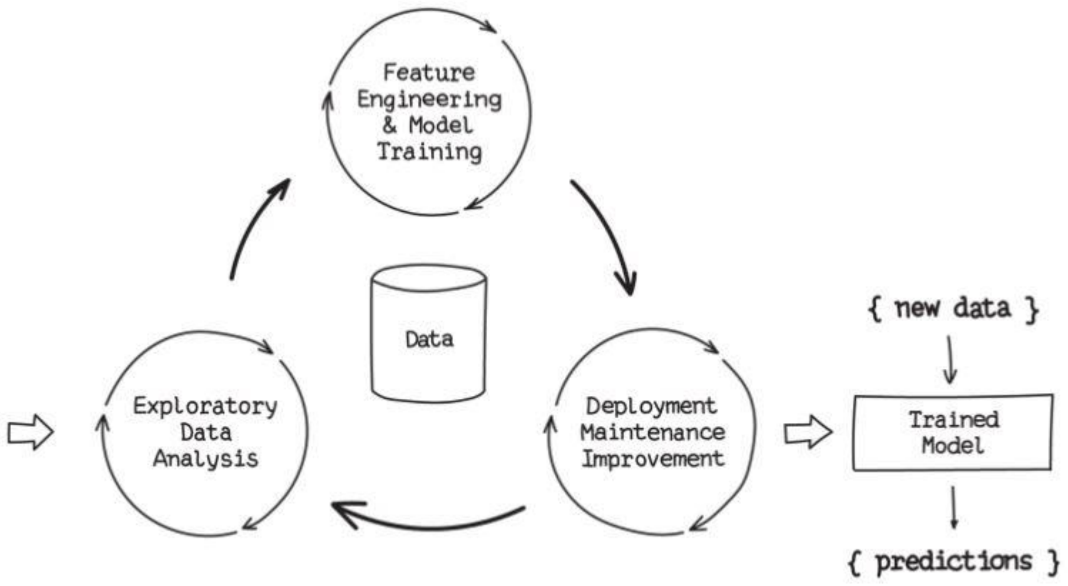

Figure 5: Process for model implementations (Dwivedi, 2021)

Conclusion

Model deployment is the last step within the objective realisation set by the business. It being the determiner of success, it can be applied in the batch model deployment and make predictions in real-time. The three main important things to be kept are the files of the model, server end script, and client end script. Computations can for predictions can be made on different machines by using the link where the server is running. There are many other platforms where you can deploy models such as AWS and Microsoft Azure. These can also be used in estimating model performance.

References

Dwivedi, R., 2021. *Hands-On-Guide To Machine Learning Model Deployment Using Flask*. [online] Analytics India Magazine. Available at: <https://analyticsindiamag.com/hands-on-guide-to-machine-learning-model-deployment-using-flask/> [Accessed 13 May 2021].

J. Demšar, .*Statistical Comparisons of Classifiers over Multiple Data Sets*, J. Mach. Learn. Res., vol. 7, pp. 1–30, 2006.

N. Japkowicz and M. Shah, *Evaluating Learning Algorithms*, 1st ed. Cambridge University Press, 2011.

Shahapurkar, S., 2016. *Crossing the Chasm: Deploying Machine Learning Analytics in Dynamic Real-World Scenarios*. Ph.D. ARIZONA STATE UNIVERSITY.

Bibliography

Brownlee, J., 2018. *How To Know if Your Machine Learning Model Has Good Performance*. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/how-to-know-if-your-machine-learning-model-has-good-performance/> [Accessed 11 May 2021].

Maila, D. and Ncube, N., 2021. *Milestone 4(Naïve Bayes)*. 1st ed. Pretoria: Belgium Campus.

Maila, D. and Ncube, N., 2021. *Milestone 5* 1st ed. Pretoria: Belgium Campus.

Maila, D. and Ncube, N., 2021. *Milestone 6* 1st ed. Pretoria: Belgium Campus.

Thompson, H., 2018. *Earnings in Certain Locations Counterbalance High Cost of Living*. [online] Esri. Available at: <https://www.esri.com/about/newsroom/publications/wherenext/salary-versus-cost-of-living/> [Accessed 6 May 2021].