We want a more accurate model to minimize the false positive.

The impact on having higher values of false positive could be dangerous for the company.

Let’s assume that the company make decisions based on the transactions predicted from the model (true positives + false positives):

1. It increases the stock of its products in order to prevent out-of-stock situation that would lead to a customer dissatisfaction;
2. It invests in more performant website in order to avoid crash or session slowness due to more traffic prediction;
3. Let’s think about the voucher approach again !!!!

The decisions could lead to a costs increase that will not be justified by the prediction. Minimizing the false positives reduces the risk of investments that would have a longer full recovery.

**Minimizing the false positives**

The GridSearch results are:

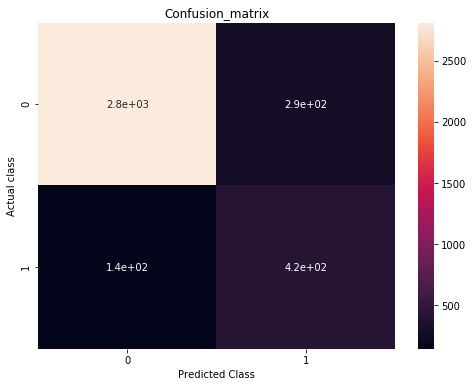
{'n\_estimators': 250}

0.9118422167709991

Random Forest with 250 trees.

|  |  |
| --- | --- |
| PageValue | 0.372773 |
| ExitRate | 0.086930 |
| ProductRelated\_Duration | 0.083587 |
| Administrative | 0.082638 |
| ProductRelated | 0.077507 |
| Administrative\_Duration | 0.066712 |
| BounceRate | 0.062556 |
| Month\_Nov | 0.032657 |
| Informational | 0.027552 |
| Informational\_Duration | 0.022183 |
| Month\_May | 0.015238 |
| Weekend | 0.010876 |
| Month\_Mar | 0.010414 |
| VisitorType\_Returning\_Visitor | 0.008499 |
| VisitorType\_New\_Visitor | 0.007804 |
| Month\_Dec | 0.006607 |
| SpecialDay | 0.005988 |
| Month\_Sep | 0.005115 |
| Month\_Oct | 0.004528 |
| Month\_Jul | 0.003846 |
| Month\_Aug | 0.003643 |
| Month\_June | 0.001810 |
| Month\_Feb | 0.000539 |

Confusion matrix using all the features available:



|  |  |
| --- | --- |
| TP: | 425 |
| TN: | 2811 |
| FP: | 294 |
| FN: | 144 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **1 feature** | **2 features** | **3 features** | **4 features** | **5 features** | **6 features** | **7 features** |
| TP: 425 | 345 | 413 | 416 | 395 | 407 | 420 | 419 |
| TN: 2811 | 2854 | 2595 | 2668 | 2729 | 2786 | 2772 | 2794 |
| FP: 294 | 251 | 510 | 437 | 376 | 319 | 333 | 311 |
| FN: 144 | 224 | 156 | 153 | 174 | 162 | 149 | 150 |
|  |  |  |  |  |  |  |  |
| **25,31%** | **39,37%** | **27,42%** | **26,89%** | **30,58%** | **28,47%** | **26,19%** | **26,36%** |

6. Optimizing the usability of the pages with higher values.

This means to analyze more deeply the behaviour of users while visiting that page.

Analysis of sales funnel in order to identify where there is the major loose of users

So for instance: monitoring time on the page, number of page visited per session

Utilize tools of Web Analytics to identify the less valuable elements of the page (such as external links)