

7.1 Logistic Regression - Predicting the Placements

Tuesday, 08 November 2022 12:00

Summary	<ul style="list-style-type: none">• Problem Statement: Given the performance of a student (input variable), how likely is the student to get placed (output variable)?																																																																																																
	<h1>Will an MBA student get placed?</h1> <p>Predicting categorical outcomes</p> <p>B-School: Business School</p>																																																																																																
	<h2>Categorical predictions</h2> <ul style="list-style-type: none">• Placement process B-Schools facilitates the graduates to pick up a job of their choice (amongst the available profiles).• The student attributes (academic performance, prior experience, internships) is expected to have a strong bearing on the outcome of the placement process.• The outcome variable in the example is binary – a student either gets a job or she doesn't!• However, the idea is more generic – the outcome variable could have several categorical values.																																																																																																
	<h2>Placement season: Data</h2> <ul style="list-style-type: none">• Let us consider the following variables that may help explain placement chances of a student:<ul style="list-style-type: none">✓ Academic performance during the undergraduate degree.✓ Academic performance during the MBA.✓ Industry experience prior to joining the MBA program.✓ Participation in the co-curricular and extra-curricular activities.• Let us assume that all these variables are scored on a scale of 10.• Lastly, our response variable is whether the student got placement or not (binary variable).																																																																																																
	<h2>Placement season: Data</h2> <ul style="list-style-type: none">• See the data in the Excel sheet. <p>Disclaimer: This is synthetic data. And it is not taken from any academic institute. This is for illustrative purposes only. This dataset should not be associated with any management program anywhere.</p>																																																																																																
	<p>This is the given data:</p> <table><tr><th>Student</th><th>MBA CGPA</th><th>Experience</th><th>UG CGPA</th><th>Extra-curricular</th><th>Day-0 placed</th></tr><tr><td>1</td><td>9.1</td><td>2.3</td><td>8.1</td><td>8.6</td><td>Placed</td></tr><tr><td>2</td><td>8.9</td><td>0</td><td>8.7</td><td>8.9</td><td>Placed</td></tr><tr><td>3</td><td>7</td><td>3.9</td><td>8</td><td>5.13</td><td>Not Placed</td></tr><tr><td>4</td><td>9.1</td><td>1.1</td><td>7.8</td><td>4.9</td><td>Not Placed</td></tr><tr><td>5</td><td>8.2</td><td>0.7</td><td>9.3</td><td>9.13</td><td>Placed</td></tr><tr><td>6</td><td>6.5</td><td>1.5</td><td>7.9</td><td>4.2</td><td>Not Placed</td></tr><tr><td>7</td><td>7.9</td><td>3.1</td><td>9.23</td><td>6.7</td><td>Not Placed</td></tr><tr><td>8</td><td>5.43</td><td>1.2</td><td>6.12</td><td>7.45</td><td>Not Placed</td></tr><tr><td>9</td><td>8.1</td><td>2.1</td><td>8.7</td><td>7.56</td><td>Placed</td></tr><tr><td>10</td><td>7.89</td><td>1.02</td><td>7.65</td><td>8.9</td><td>Placed</td></tr><tr><td>11</td><td>8.65</td><td>2.3</td><td>7.98</td><td>5.1</td><td>Placed</td></tr><tr><td>12</td><td>9.45</td><td>2.5</td><td>9.2</td><td>6.8</td><td>Placed</td></tr><tr><td>13</td><td>7.8</td><td>5.1</td><td>8.9</td><td>9.1</td><td>Placed</td></tr><tr><td>14</td><td>9.01</td><td>0.3</td><td>9.2</td><td>7.8</td><td>Placed</td></tr><tr><td>15</td><td>6.8</td><td>2.3</td><td>9.8</td><td>9.1</td><td>Not Placed</td></tr></table>	Student	MBA CGPA	Experience	UG CGPA	Extra-curricular	Day-0 placed	1	9.1	2.3	8.1	8.6	Placed	2	8.9	0	8.7	8.9	Placed	3	7	3.9	8	5.13	Not Placed	4	9.1	1.1	7.8	4.9	Not Placed	5	8.2	0.7	9.3	9.13	Placed	6	6.5	1.5	7.9	4.2	Not Placed	7	7.9	3.1	9.23	6.7	Not Placed	8	5.43	1.2	6.12	7.45	Not Placed	9	8.1	2.1	8.7	7.56	Placed	10	7.89	1.02	7.65	8.9	Placed	11	8.65	2.3	7.98	5.1	Placed	12	9.45	2.5	9.2	6.8	Placed	13	7.8	5.1	8.9	9.1	Placed	14	9.01	0.3	9.2	7.8	Placed	15	6.8	2.3	9.8	9.1	Not Placed
Student	MBA CGPA	Experience	UG CGPA	Extra-curricular	Day-0 placed																																																																																												
1	9.1	2.3	8.1	8.6	Placed																																																																																												
2	8.9	0	8.7	8.9	Placed																																																																																												
3	7	3.9	8	5.13	Not Placed																																																																																												
4	9.1	1.1	7.8	4.9	Not Placed																																																																																												
5	8.2	0.7	9.3	9.13	Placed																																																																																												
6	6.5	1.5	7.9	4.2	Not Placed																																																																																												
7	7.9	3.1	9.23	6.7	Not Placed																																																																																												
8	5.43	1.2	6.12	7.45	Not Placed																																																																																												
9	8.1	2.1	8.7	7.56	Placed																																																																																												
10	7.89	1.02	7.65	8.9	Placed																																																																																												
11	8.65	2.3	7.98	5.1	Placed																																																																																												
12	9.45	2.5	9.2	6.8	Placed																																																																																												
13	7.8	5.1	8.9	9.1	Placed																																																																																												
14	9.01	0.3	9.2	7.8	Placed																																																																																												
15	6.8	2.3	9.8	9.1	Not Placed																																																																																												

16	7.14	0.4	8.56	6.89	Not Placed
17	8.56	1.3	8.34	7.65	Not Placed
18	7.4	0	8.7	5.6	Not Placed
19	8.23	0	7.8	6.5	Not Placed
20	7.3	0	7.1	8.56	Not Placed
21	9.1	0	9.8	8.9	Placed
22	7.8	1.3	6.7	8.2	Not Placed
23	5.6	2.3	8.7	6.9	Not Placed
24	8.1	1.78	7.28	7.9	Not Placed
25	8.79	3.98	9.11	7	Placed
26	7.19	0.2	8.16	6.7	Not Placed
27	8.18	1.9	8.25	8.9	Not Placed

- We've coded the Day-0 placed column as:
 - Placed: 1
 - Not Placed: 0

Placement season: The Ask

- Given the data, what are we interested in?
- Of course, try and build a model that can help predict the chances of a student for whom the input data is available.
- Towards that, let us define the variables as:

X_1 = Academic performance during the undergraduate degree.

X_2 = Academic performance during the MBA.

X_3 = Industry experience prior to joining the MBA program.

X_4 = Participation in the co-curricular and extra-curricular activities.

$Y = 1$ if the student gets placed, and zero otherwise.