# Tutorial Business Analytics

Homework 5 - Solution

**Homework 5.1**

The owner of a restaurant examines the behavior of his clients while they wait for a table. She uses a Naïve Bayes classifier and the following set of training data.

| Alternative choice | Raining | Waiting time | Stay or go |
|---|---|---|---|
| Yes | No | 0 – 10 minutes | Stay |
| No | Yes | 11 – 20 minutes | Stay |
| Yes | No | > 20 minutes | Go |
| No | No | 0 – 10 minutes | Stay |
| No | Yes | 0 – 10 minutes | Stay |
| Yes | Yes | 11 – 20 minutes | Go |
| Yes | Yes | 11 – 20 minutes | Go |
| Yes | No | 0 – 10 minutes | Stay |
| No | No | 0 – 10 minutes | Stay |
| Yes | No | 11 – 20 minutes | Go |
| Yes | Yes | 11 – 20 minutes | Stay |
| No | Yes | > 20 minutes | Stay |

Calculate the probability for every class and classify below-mentioned instance as stay or go.

| Alternative choice | Raining | Waiting time | Stay or go |
|---|---|---|---|
| Yes | No | 0 – 10 minutes | ? |

**Solution**

Frequency tables:

| Alternative choice | Stay | Go |
|---|---|---|
| Yes | 3 | 4 |
| No | 5 | 0 |

| Raining | Stay | Go |
|---|---|---|
| Yes | 4 | 2 |
| No | 4 | 2 |

| Waiting | Stay | Go |
|---|---|---|
| 0 – 10 minutes | 5 | 0 |
| 11 – 20 minutes | 2 | 3 |
| > 20 minutes | 1 | 1 |

| Stay or go | |
|---|---|
| Stay | 8 |
| Go | 4 |

Likelihood tables (w/o +1):

| Alternative choice | Stay | Go |
|---|---|---|
| Yes | 3/8 | 4/4 |
| No | 5/8 | 0/4 |

| Raining | Stay | Go |
|---|---|---|
| Yes | 4/8 | 2/4 |
| No | 4/8 | 2/4 |

| Waiting | Stay | Go |
|---|---|---|
| 0 – 10 minutes | 5/8 | 0/4 |
| 11 – 20 minutes | 2/8 | 3/4 |
| > 20 minutes | 1/8 | 1/4 |

| Stay or go | |
|---|---|
| Stay | 8/12 |
| Go | 4/12 |

With zero-frequency increment (+1)

| Alternative choice | Raining | Waiting | Stay | Go | Stay | Go |
|---|---|---|---|---|---|---|
| Yes | No | 0 – 10 | 0.07273 | 0.01984 | 0.786 | 0.214 |

Without zero-frequency increment

| Alternative choice | Raining | Waiting | Stay | Go | Stay | Go |
|---|---|---|---|---|---|---|
| Yes | No | 0 – 10 | 0.07813 | 0 | 1 | 0 |

**Homework 5.2**

The following table contains empirical data about car ownership.

| Age | Annual salary | Children | Car |
|---|---|---|---|
| < 30 | < 35,000 | No | No |
| < 30 | < 35,000 | Yes | No |
| < 30 | ≥ 35,000 | Yes | No |
| < 30 | < 35,000 | No | Yes |
| ≥ 30 | < 35,000 | No | Yes |
| ≥ 30 | < 35,000 | Yes | No |
| ≥ 30 | ≥ 35,000 | No | Yes |
| ≥ 30 | < 35,000 | Yes | Yes |
| ≥ 30 | ≥ 35,000 | Yes | Yes |

a) Calculate the rule set using the 1-rule classification.

b) Use the obtained rule set to determine the classification for below-mentioned data record.

| Age | Annual salary | Children | Car |
|---|---|---|---|
| ≥ 30 | ≥ 35,000 | No | ? |

**Solution**

a)

| Attribute | Rule | Error rate | Total error rate |
|---|---|---|---|
| Age | < 30 → no car<br>≥ 30 → car | 1/4<br>1/5 | 2/9 |
| Annual salary | < 35,000 → no car<br>≥ 35,000 → car | 3/6<br>1/3 | 4/9 |
| Children | Yes → no car<br>No 30 → car | 2/5<br>1/4 | 3/9 |

b) Age ≥ 30 → Car