HW1-Data622 - Qn1

Below is the dataset of the prospecting customer printed in a tabular format.
Given below are the calculation done for finding the prior probabilities and the conditional probability using simple counting.

age.group	networth	status	credit_rating	classprospect
youth	high	employed	fair	no
youth	high	employed	excellent	no
middle	high	employed	fair	yes
senior	medium	employed	fair	yes
senior	low	unemployed	fair	yes
senior	low	unemployed	excellent	no
middle	low	unemployed	excellent	yes
youth	medium	employed	fair	no
youth	low	unemployed	fair	yes
senior	medium	unemployed	fair	yes
youth	medium	unemployed	excellent	yes
middle	medium	employed	excellent	yes
middle	high	unemployed	fair	yes
senior	medium	employed	excellent	no

1)Compute prior probabilities for the Prospect Yes/No

Prior Priorities P(prospect=yes) = 9/14 = 0.64
Prior Priorities P(prospect=no) = 5/14 = 0.36

2) Compute the conditional probabilities
P(age-group=youth|prospect=yes) and
P(age-group=youth|prospect=no)
where age-group is a predictor variable.Compute the conditional probabilities for each predictor variable,namely,(age_group,networth,status,credit_rating)
Conditional Probabilities:

P(age-group=youth|prospect=yes) = 2/9 = 0.22
P(age-group=middle|prospect=yes) = 4/9 = 0.44
P(age-group=senior|prospect=yes) = 3/9 = 0.33
P(age-group=youth|prospect=no) = 3/5 = 0.6
P(age-group=middle|prospect=no) = 0/5 = 0
P(age-group=senior|prospect=no) = 2/5 = 0.4

P(networth=high|prospect=yes) = 2/9 = 0.22
P(networth=low|prospect=yes) = 3/9 = 0.33
P(networth=medium|prospect=yes) = 4/9 = 0.44
P(networth=high|prospect=no) = 2/5 = 0.4
P(networth=low|prospect=no) = 1/5 = 0.2
P(networth=medium|prospect=no) = 2/5 = 0.4

P(status=employed|prospect=yes) = 3/9 = 0.33
P(status=employed|prospect=no) = 4/5 = 0.8
P(status=unemployed|prospect=yes) = 6/9 = 0.67
P(status=unemployed|prospect=no) = 1/5 = 0.2

P(credit=fair|prospect=yes) = 6/9 = 0.67
P(credit=fair|prospect=no) =2/5 = 0.4
P(credit=excellent|prospect=yes) = 3/9 = 0.33
P(credit=excellent|prospect=no) = 3/5 = 0.6

3) Assuming the assumptions of Naive Bayes are met, compute the posterior probability P(prospect|X) where X is one of the predictor variables.

P(prospect=no|age-group,networth,status,credit)
P(prospect=yes|age-group,networth,status,credit)

1.P(YES|youth,medium,unemployed,fair)

= P(Yes) * P(youth|Yes) * P(medum| Yes) * P(unemployed| Yes) * P(fair| Yes)

= 0.64 * 0.22 * 0.44 * 0.67 * 0.67 = **0.03**

2. P(YES| youth,low,unemployed,excellent)

= P(Yes) * P(youth|Yes) * P(low| Yes) * P(unemployed| Yes) * P(excellent| Yes)

= 0.64 * 0.22 * 0.33 * 0.67 * 0.33 = **0.01**

P(YES| senior,medium,unemployed,excellent)

= P(Yes) * P(senior|Yes) * P(medium| Yes) * P(unemployed| Yes) * P(excellent| Yes)

= 0.64 * 0.33 * 0.44 * 0.67 * 0.33 = **0.02**

P(YES| middle,medium,unemployed,fair)

= P(Yes) * P(middle|Yes) * P(medium| Yes) * P(unemployed| Yes) * P(fair| Yes)

= 0.64 * 0.44 * 0.44 * 0.67 * 0.67 = **0.06**

P(YES| senior,low,employed,fair,yes)

= P(Yes) * P(senior|Yes) * P(low| Yes) * P(employed| Yes) * P(fair| Yes)

= 0.64 * 0.33 * 0.33 * 0.33 * 0.67 = **0.02**

P(YES| senior,low,employed,excellent,no)

= P(Yes) * P(senior|Yes) * P(low| Yes) * P(employed| Yes) * P(excellent| Yes)

= 0.64 * 0.33 * 0.33 * 0.33 * 0.33 = **0.008**

P(NO|youth,medium,unemployed,fair)

= P(No) * P(youth|No) * P(medium| No) * P(unemployed| No) * P(fair| No)

= 0.36 * 0.6 * 0.4 * 0.2 * 0.4 = **0.007**

P(NO| youth,low,unemployed,excellent)

= P(No) * P(youth|No) * P(low| No) * P(unemployed| No) * P(excellent| No)

= 0.36 * 0.6 * 0.2 * 0.2 * 0.6 = **0.005**

P(NO| senior,medium,unemployed,excellent)

= P(No) * P(senior|No) * P(medium| No) * P(unemployed| No) * P(excellent| No)

= 0.36 * 0.4 * 0.4 * 0.2 * 0.6 = **0.007**

P(NO| middle,medium,unemployed,fair)

= P(No) * P(middle|No) * P(medium| No) * P(unemployed| No) * P(fair| No)

= 0.36 * 0 * 0.4 * 0.2 * 0.4 = **0**

P(NO| senior,low,employed,fair,yes)

= P(No) * P(senior|No) * P(low| No) * P(employed| No) * P(fair| No)

= 0.36 * 0.4 * 0.2 * 0.8 * 0.4 = **0.009**

P(NO| senior,low,employed,excellent,no)

= P(No) * P(senior|No) * P(low| No) * P(employed| No) * P(excellent| No)

= 0.36 * 0.4 * 0.2 * 0.8 * 0.6 = **0.01**