Assignment C2: BAI Tool For Profit Maximization

Roll No.4431

1 Problem Statement:

Select and Industrial Sector and write a BAI tool for maximizing the profit.(Optional: Placement Companies)

2 Learning Objectives:

1. Understand the usage of Business Analytics in Industrial Sector.
2. Develop a tool for analysing the student data and recommending the company for placement.

3 Learning Outcomes:

1. Implementation of BAI tool to recommend companies for placements to students.
2. Ability to develop applications for BAI.

4 Theory:

4.1 Introduction:

In business analytics, big data needs to be analysed to identify patterns that enables a business to structure itself so as to get maximum profits. Various profit maximization approaches can be followed. Two have been used in this case:

A sample problem in this case is considered as:
Placement data is available per student that includes information such as Percentage, Number of projects, Internships, Marks, Salary etc. The available data is analysed so as to obtain the pattern that provides the maximum Salary.

Alternatively, given parameters such as Percentage, Marks etc., the system identifies the possible placement scenarios: e.g, Company, Salary.

Two approaches have been used:
1. Simple cumulative value calculator
2. Support Vector Machine

Simple cumulative value calculator:
In this approach, the scores of each individual parameter is added, and a cumulative value is obtained. For the maximum salary, the corresponding parameters are displayed. When parameters are given as input, the cumulative value of those parameters is compared with the available dataset. If a nearby match is found, the corresponding company name and salary is displayed.

Support Vector Machine:
In this approach, an SVM is used to determine the probability of getting placed in a company with salary greater than certain value. SVM takes in input as the dataset, and then processes the input parameters to provide the best possible probability of the person getting placed in the companies.

R programming language is used for this system.

4.2 Support Vector Machines:

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

4.3 R Programming Language:

R is a programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. Polls, surveys of data miners, and studies of scholarly literature databases show that R's popularity has increased substantially in recent years.

R is an implementation of the S programming language combined with lexical scoping semantics inspired by Scheme. S was created by John Chambers while at Bell Labs. There are some important differences, but much of the code written for S runs unaltered.

4.4 Code:

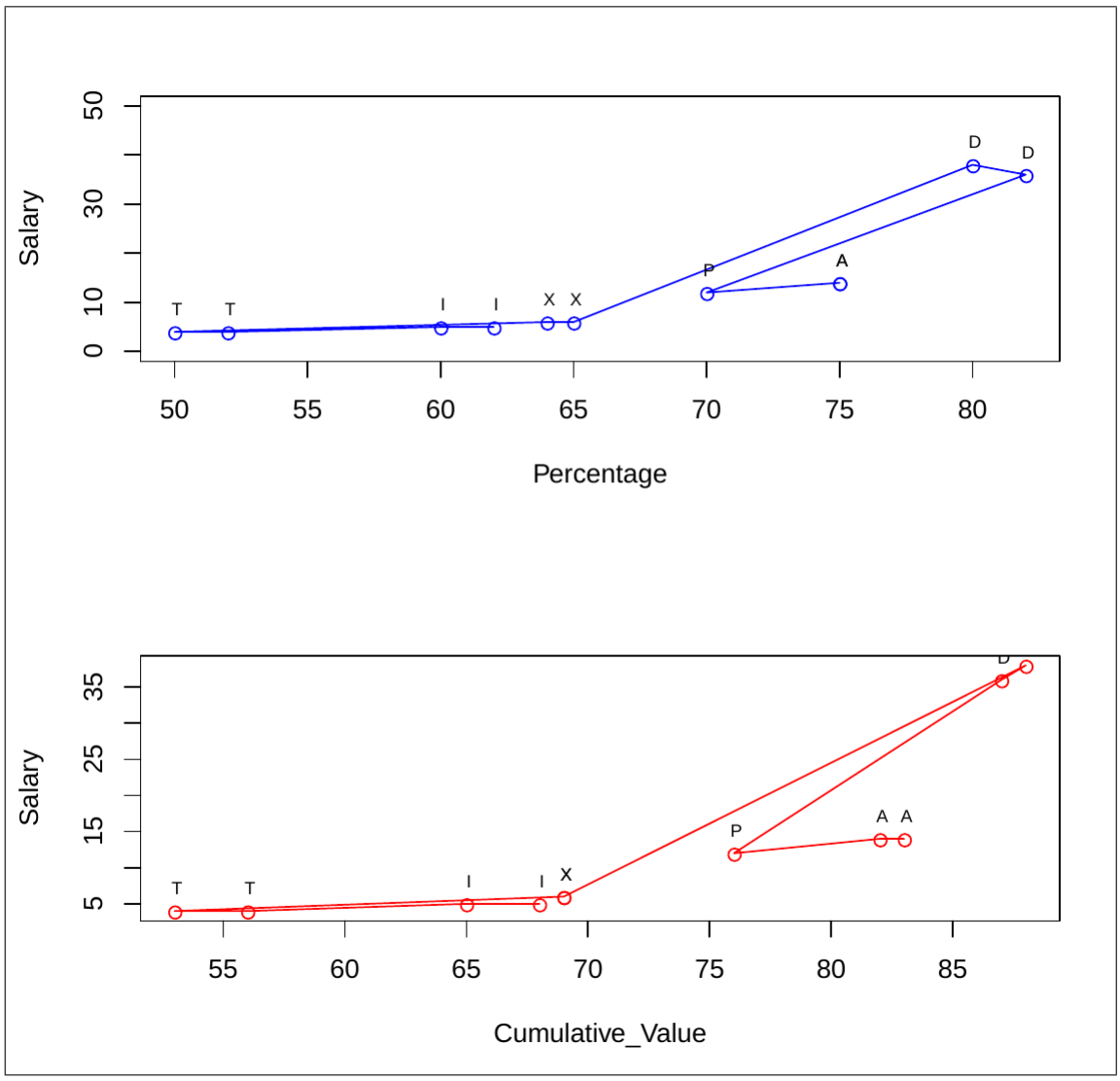
1. Approach based on Cumulative Value Calculation:

```
//R Script:
cat("Reading placement data from records.csv file.Data is:")
cat("\n")
heads<-read.csv("records.csv")
heads
Percentage=heads$Percentage
Projects=heads$Projects
Internships=heads$Internships
Papers=heads$Papers
Company=heads$Company
Salary=heads$Salary
#curve(x^2+x,from=0, to=100)
calculate<-function(Percentage,Projects,Internships,Papers){
  return(Percentage+Projects+Internships+Papers)
}
Cumulative_Value=calculate(Percentage,Projects,Internships,Papers)
par(mfrow = c(2,1))
plot(Percentage, Salary,type="o",col="blue",ylim = c(0,50))
text(Percentage, Salary, labels=Company, cex= 0.7, pos=3)

plot(Cumulative_Value,Salary,type="o",col="red")
text(Cumulative_Value,Salary, labels=Company, cex= 0.7, pos=3)

max_value=max(Salary)
index=match(max_value,Salary)
cat("The maximum profit (highest salary) is in the case of: ")
cat("\n")
cat(paste("Percentage: ",Percentage[index]))
cat("\n")
cat(paste("Projects: ",Projects[index]))
cat("\n")
cat(paste("Internships: ",Internships[index]))
cat("\n")
cat(paste("Papers: ",Papers[index]))
cat("\n")
```

```
cat(paste("With the company being: ",Company[index]))
cat("\n")
cat(paste("Cumulative Value being: ",Cumulative_Value[index]))
cat("\n")
cat("**** Refer Rplots.pdf for overview****")
cat("\n")
cat("Enter data to check possible placement company and salary:")
cat("\n")
cat("Enter percentage: ")
percent1 <- readLines("stdin",n = 1)
cat("Enter number of projects: ")
projects1 <- readLines("stdin",n=1)
cat("Enter number of internships: ")
internships1 <- readLines("stdin",n=1)
cat("Enter number of papers: ")
papers1 <- readLines("stdin",n=1)

percent1<-as.integer(percent1)
projects1<-as.integer(projects1)
internships1<-as.integer(internships1)
papers1<-as.integer(papers1)

check_value=percent1+projects1+internships1+papers1
temp<-c((check_value-1):(check_value+1))
store_index=-1
for (i in 1:length(temp)){
  for (j in 1:length(Cumulative_Value)){
    if(temp[i]==Cumulative_Value[j]){
        store_index=j
        break
    }
  }
}
if(store_index!=-1){
  print(paste("Recommended company is:",Company[store_index]," with salary: ",Sa
}else{
  print("Cannot determine possible placement scenario")
}

//records.csv (Dataset):
Percentage, Projects, Internships, Papers, Company, Salary
75, 5, 2, 1, A, 14
75, 4, 0, 3, A, 14
70, 4, 1, 1, P, 12
82, 5, 0, 0, D, 36
80, 6, 1, 1, D, 38
```

65, 4, 0, 0, X, 6
64, 4, 1, 0, X, 6
50, 3, 0, 0, T, 4
52, 3, 1, 0, T, 4
60, 4, 1, 0, I, 5
62, 5, 0, 1, I, 5

4.5 Output using approach 1:



```
[nsk@ARCH_NSK C2]$ Rscript C2.r
Reading placement data from records.csv file.Data is:
   Percentage Projects Internships Papers Company Salary
1          75        5           2      1       A     14
2          75        4           0      3       A     14
3          70        4           1      1       P     12
4          82        5           0      0       D     36
5          80        6           1      1       D     38
6          65        4           0      0       X      6
7          64        4           1      0       X      6
8          50        3           0      0       T      4
9          52        3           1      0       T      4
10         60        4           1      0       I      5
11         62        5           0      1       I      5
The maximum profit (highest salary) is in the case of:
Percentage:  80
Projects:  6
Internships:  1
Papers:  1
With the company being:   D
Cumulative Value being:  88
**** Refer Rplots.pdf for overview****
Enter data to check possible placement company and salary:
Enter percentage: 70
Enter number of projects: 4
Enter number of internships: 1
Enter number of papers: 1
[1] "Recommended company is:  P  with salary:  12"
```

4.6 Output using approach 2:

```
#Code: (Interactive Mode)

> library('e1071')
> #reading dataset of placement record
> data <- read.csv("placement-data.csv")
> #modeling the input data-set with svm
> model <- svm(data[,-ncol(data)], data[,ncol(data)], gamma=10)
> #predicting the output for twenty samples
> print(predict(model, data[c(1:20),-ncol(data)]))
```

```
          1           2           3           4           5
6           7
0.05029538 0.05048614 0.94974144 0.94981699 0.05028727 0.94954545 0.94954412
          8           9          10          11          12
13          14
0.94974263 0.94974141 0.94954545 0.05029538 0.05056266 0.05029022 0.94942183
         15          16          17          18          19
20
0.94954254 0.94955249 0.05048509 0.05056268 0.05049376 0.05028876
> #here values which are closed to 1 means those
students can be placed in a company
#placement-data.csv
> read
    X10.th X12.th Fe1.st.sem Fe.2.ndb.sem SE.1.st.sem SE.2.nd.Sem TE.1.se.Sem
1    85.46  80.17        8.29          8.93         8.92
7.39         7.45
2    87.46  77.17        8.25          8.11         8.42
8.70         7.32
3    74.93  71.17        8.06          7.39         7.08
7.48         7.05
4    78.53  84.00        8.98          8.41         8.58
7.78         7.91
5    72.66  77.67        7.94          6.89         6.67
5.96         6.05
6    84.66  82.33        8.06          7.76         7.63
8.30         8.27
7    87.86  76.50        7.77          7.78         7.79
7.30         7.59
8    83.00  78.80        9.13          8.52         8.38
8.87         8.50
9    80.53  83.17        8.92          8.50         9.17
9.09         8.73
10   85.60  81.00        9.29          9.13         8.88
8.57         8.55
11   72.93  79.50        7.56          7.22         7.29
6.74         6.41
12   82.40  80.00        7.73          7.11         7.00
6.52         6.86
13   79.46  80.17        8.10          8.57         8.42
8.57         8.00
14   83.73  83.33        8.35          8.02         7.38
7.91         7.91
15   78.13  73.50        7.58          6.52         6.38
6.00         6.36
16   78.40  75.00        6.50          5.72         6.00
5.74         6.05
```

17	80.13	82.66	7.83	7.89	7.79	7.65	7.68
18	64.26	57.83	7.00	6.43	5.63	5.74	5.50
19	77.46	79.83	7.50	7.85	6.92	6.74	6.68
20	82.40	70.17	8.06	7.09	6.79	7.48	6.68
21	79.86	81.50	8.46	7.96	7.79	7.61	7.55
22	79.73	62.67	7.98	7.50	5.83	6.00	5.50
23	88.26	88.00	7.65	7.54	6.33	5.83	6.18
24	84.40	89.33	8.83	8.09	7.42	7.96	8.09
25	84.00	70.00	8.06	7.76	7.54	7.70	6.41
26	87.46	85.33	8.69	8.52	9.38	9.13	8.45
27	74.00	88.00	7.94	7.39	6.96	6.26	5.55
28	83.33	79.67	8.81	8.70	8.13	7.35	7.09
29	70.13	77.17	8.40	8.24	7.75	7.52	7.36
30	82.28	75.00	7.54	7.28	6.79	6.74	6.27
31	82.40	83.50	8.29	8.41	7.46	7.39	6.32
32	84.80	67.83	6.94	7.22	6.29	6.04	5.82
33	82.93	85.00	8.31	8.63	8.83	8.39	7.27
34	73.47	77.00	7.73	6.87	6.42	6.26	6.36
35	78.53	66.83	7.06	7.59	6.17	6.17	5.77
36	73.73	77.83	7.71	7.24	6.63	7.26	6.82
37	78.93	67.67	6.83	6.09	5.96	6.13	5.32
38	83.00	67.00	7.02	7.54	6.75	5.87	5.14
39	79.86	78.16	8.06	7.24	6.71	6.00	5.77

40	74.00	65.50	7.10	7.93	7.58	7.57	6.95
41	86.00	81.67	8.23	8.02	7.50	7.78	7.59
42	73.60	78.33	7.73	7.50	6.21	6.57	6.27
43	90.13	83.50	8.40	8.20	7.71	7.65	8.09
44	65.20	66.67	6.29	5.98	6.29	5.83	5.86
45	77.73	67.83	7.85	8.09	7.42	6.91	6.95
46	80.80	71.17	7.71	7.22	6.79	6.35	6.23
47	75.33	58.67	6.73	6.83	6.88	6.70	7.05
48	80.40	62.40	8.13	7.54	6.46	6.70	6.18
49	78.13	69.83	8.15	8.15	7.21	6.57	5.41
50	86.00	84.83	7.75	7.63	7.00	7.00	6.68
51	82.13	80.17	8.08	7.59	7.63	7.70	7.36
52	84.80	84.17	8.06	8.11	7.63	7.35	6.82
53	83.06	80.05	6.98	6.96	6.13	6.13	5.95
54	78.53	72.00	7.96	7.57	7.54	7.13	6.23
55	80.00	82.17	7.96	7.96	8.67	8.48	8.14
56	81.46	85.50	8.71	9.02	8.58	9.13	8.59
57	80.93	77.75	6.70	6.80	6.50	6.90	6.70
58	60.66	60.00	7.23	7.15	7.25	7.52	7.09
59	77.33	81.00	8.56	8.39	8.29	8.65	8.55
60	85.73	88.67	8.46	8.83	6.92	7.22	7.73
61	83.20	75.52	7.21	7.56	7.25	7.52	7.36
62	71.06	67.50	7.67	7.23	6.92	7.43	6.86

63	56.66	60.25	6.25	5.50	6.67
6.96		6.77			
64	69.46	62.43	5.70	6.10	8.17
7.87		8.00			
65	85.73	88.67	8.46	8.83	6.92
7.22		7.73			
66	85.73	88.67	8.46	8.83	6.92
7.22		7.73			
67	79.46	61.00	7.96	6.80	5.96
6.08		6.68			

	TE.2.nd.Sem	Active.Backlog	Passive.Backlog	Placed
1	6.88	0	0	0
2	6.38	0	0	0
3	6.17	0	0	1
4	7.75	0	0	1
5	6.38	0	0	0
6	7.58	0	0	1
7	7.46	0	0	1
8	8.50	0	0	1
9	8.46	0	0	1
10	7.96	0	0	1
11	6.50	0	0	0
12	6.92	0	0	0
13	7.63	0	0	0
14	7.17	0	0	1
15	7.17	0	1	1
16	6.46	0	0	1
17	7.96	0	0	0
18	6.33	0	0	0
19	7.08	0	0	0
20	6.46	0	0	0
21	7.04	0	0	0
22	5.23	0	1	0
23	6.78	0	1	0
24	7.96	0	0	1
25	6.96	0	0	0
26	7.75	0	0	1
27	5.35	0	1	0
28	7.63	0	0	1
29	7.83	0	0	1
30	6.56	0	1	0
31	6.79	0	0	0
32	6.25	0	0	0
33	7.00	0	0	0
34	6.04	0	0	1
35	6.50	0	0	1

36	6.88	0	0	1
37	6.50	2	0	1
38	8.25	0	1	1
39	6.21	0	0	0
40	6.58	0	0	0
41	7.50	0	0	1
42	6.13	0	0	0
43	7.71	0	0	0
44	6.23	1	1	1
45	6.46	0	0	1
46	6.75	0	0	1
47	6.83	0	0	1
48	6.29	0	0	0
49	5.92	0	0	0
50	7.29	0	0	1
51	7.04	0	0	0
52	6.96	0	0	1
53	5.24	0	2	0
54	6.50	0	0	1
55	7.25	0	0	1
56	8.79	0	0	1
57	6.70	0	0	0
58	7.58	0	0	0
59	7.58	0	0	1
60	7.96	0	0	1
61	6.88	0	0	1
62	6.04	0	0	0
63	6.83	0	0	0
64	7.13	0	0	0
65	7.96	0	0	1
66	7.96	0	0	1
67	6.50	1	0	

5 Conclusion:

Thus a BAI tool was successfully developed to maximize profit based on salary in a dataset that contained placement data.