Impact of Artificial intelligence on Industries

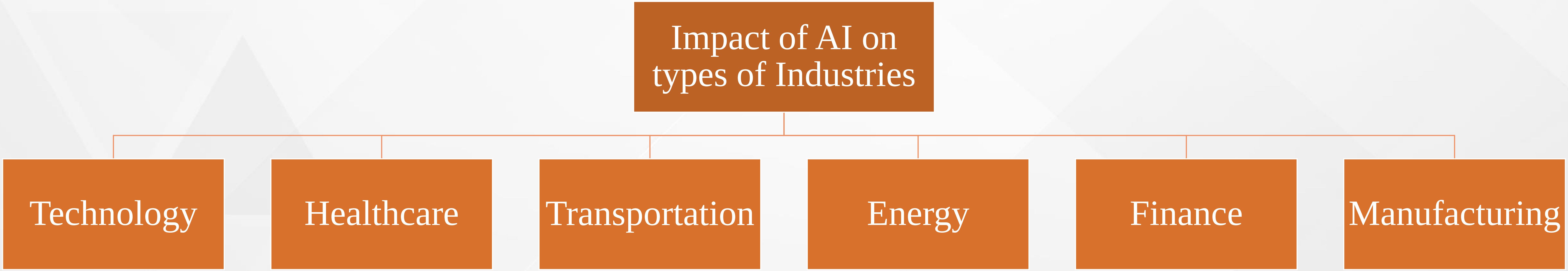
Team comprises of:

- Abhishek Kashid
- Gaurav Humane
- Harsh Mandare
- Saurabh Tripathi
- Vinayak Hapte

Project Mentor: DR. V.U. Dixit

Introduction

AI, or artificial intelligence, is the development of computer systems that can perform tasks that typically require human intelligence, such as learning, problem-solving, and decision- making. It's like giving computers a brain to think and act like people do. AI uses data and rules to figure out patterns and make predictions or decisions.



Advantages of AI

- Efficiency
- Automation
- Cost saving
- Safety
- Accuracy
- Accessibility

Disadvantages of AI

- Job Displacement
- Dependence
- Regulatory challenges
- Privacy concerns
- Bias
- Lack of transparency

Objectives

Following are the objectives of our survey :

1. To study the impact of Artificial Intelligence on the Efficiency of the work in the Industries.
2. To analyze the impact of Artificial Intelligence on the Expenditure in the Industries.
3. To study the impact of Artificial Intelligence on Customer Satisfaction in Industries.

Research Methodology

Steps involved in conducting the survey:

- Defining our objective and scope of the survey.
- Literature survey -Review existing studies and theories related to AI and its impact.
- Develop specific questions and hypotheses that our research will address.
- Pilot Survey.
- Modifying questionnaires.
- Data Collection (Used Primary data).
- Data coding and Data entry.
- Data analysis.
- Conclusion.
- Preparation of Project Report.

Methodology : We conducted a pilot survey of sample size 52.After which we made the necessary changes in our Questionnaire .We then surveyed 346 individually and via Google forms with the help of Non-probability Convenient Sampling. We cleaned the data after entering it in the MS-EXCEL. After data cleaning ,we were left with final sample size 292.

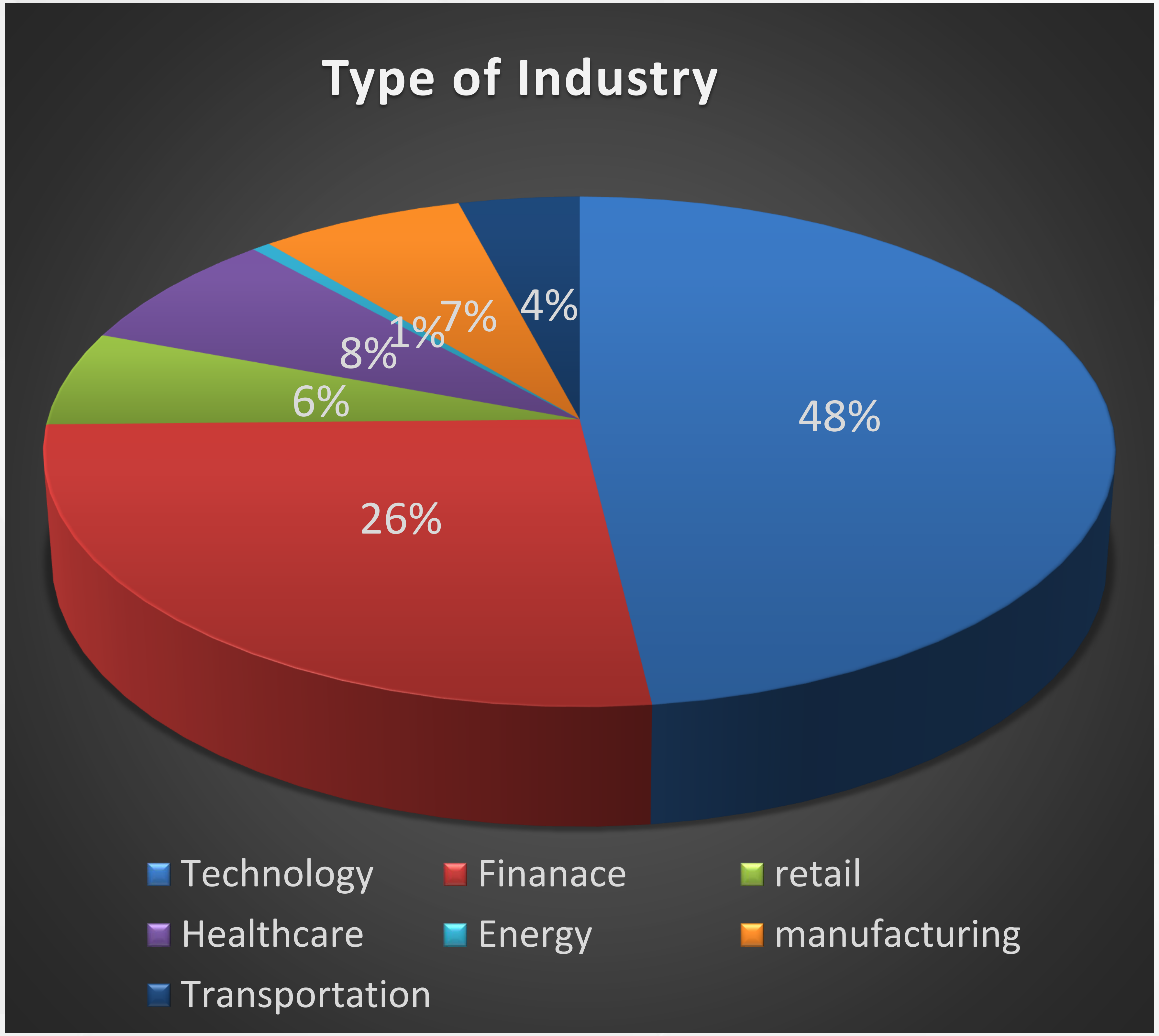
DATA CLEANING :

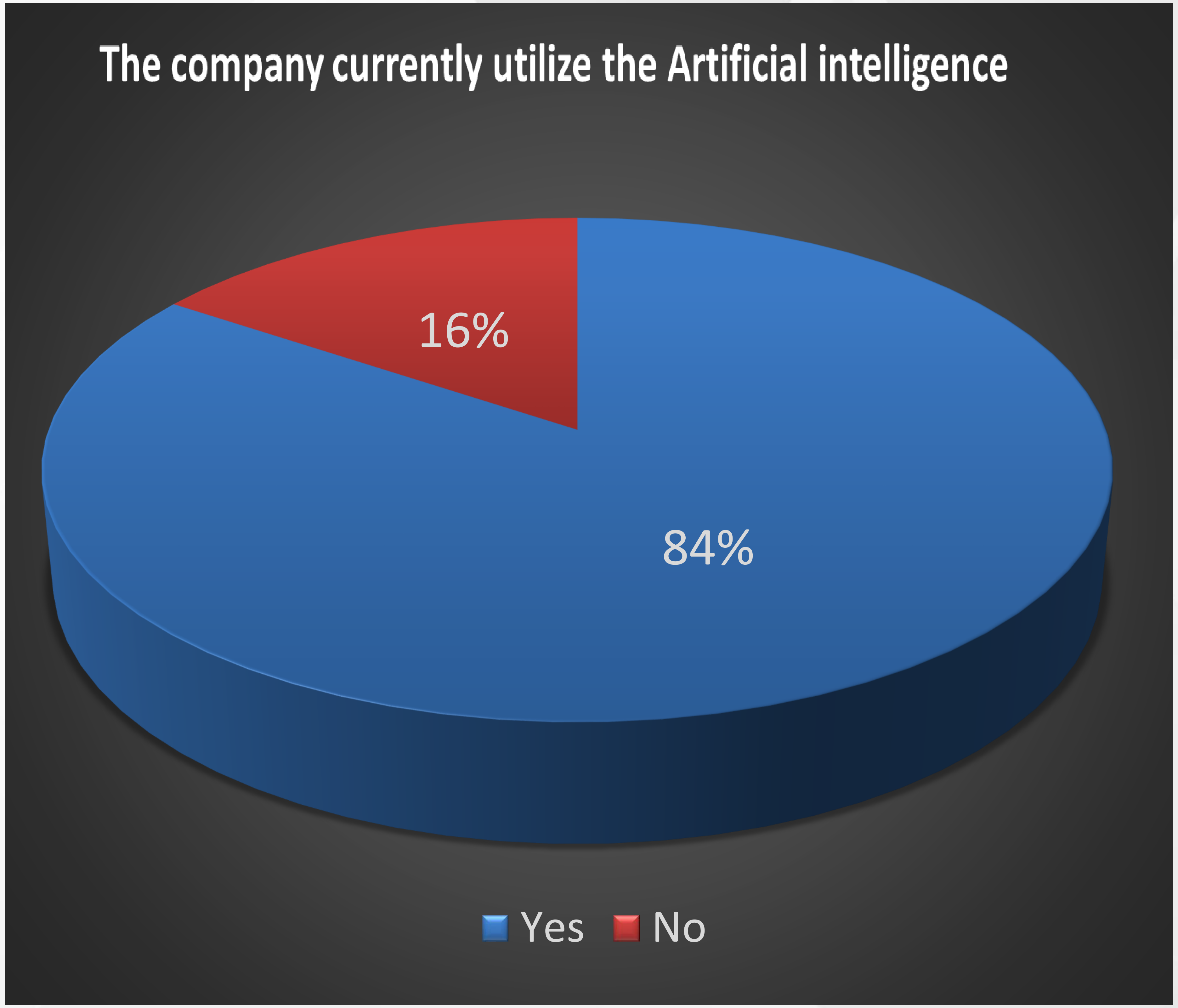
Removal of unwanted observations	**Handling missing data**
Managing unwanted outliers	**Fixing structural errors**

Techniques used

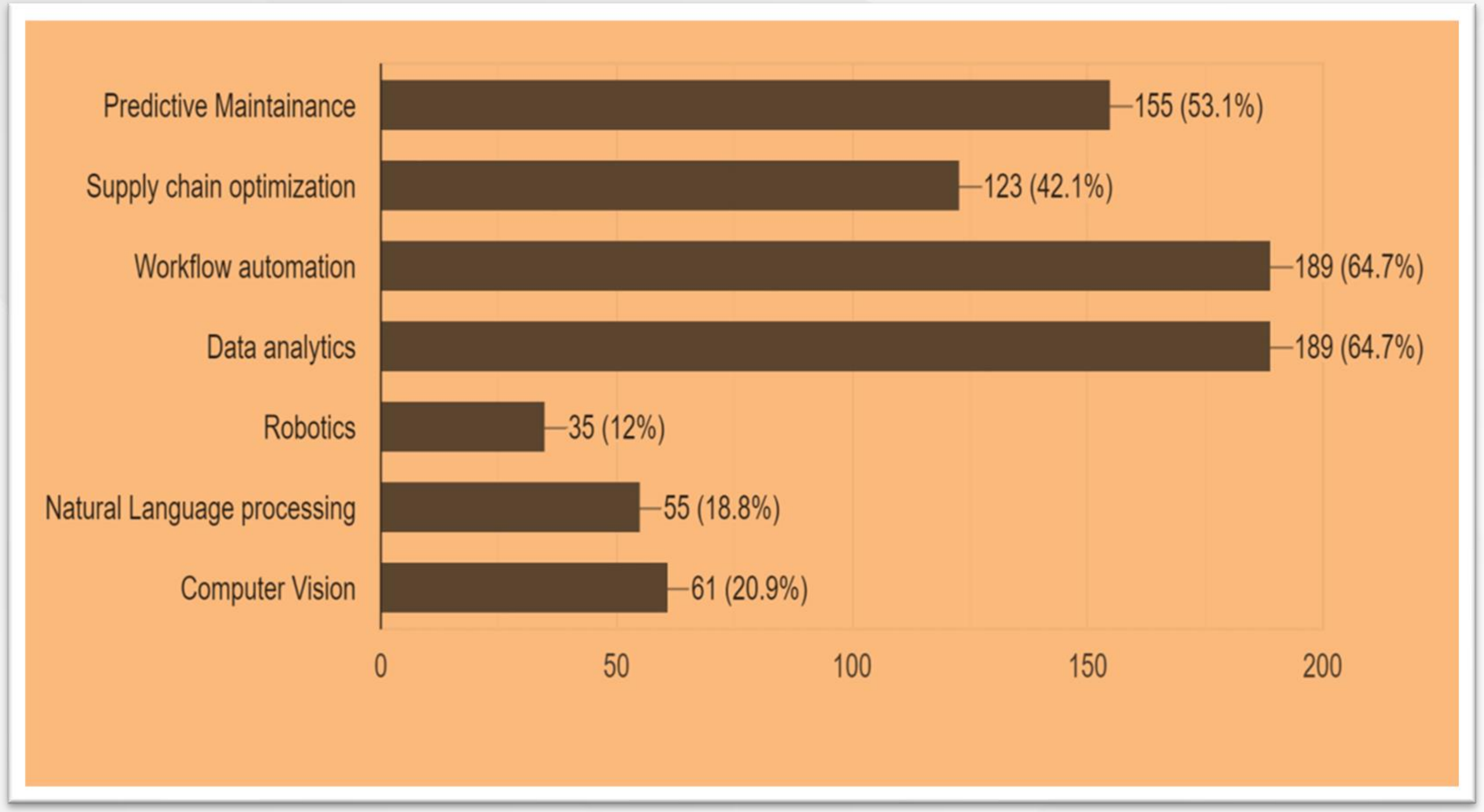
- ❖Graphical Representation
- ❖Ordinal Logistic Regression
- ❖Confusion matrix
- ❖Likelihood ratio test
- ❖Ordinal Forest technique

Exploratory data analysis:

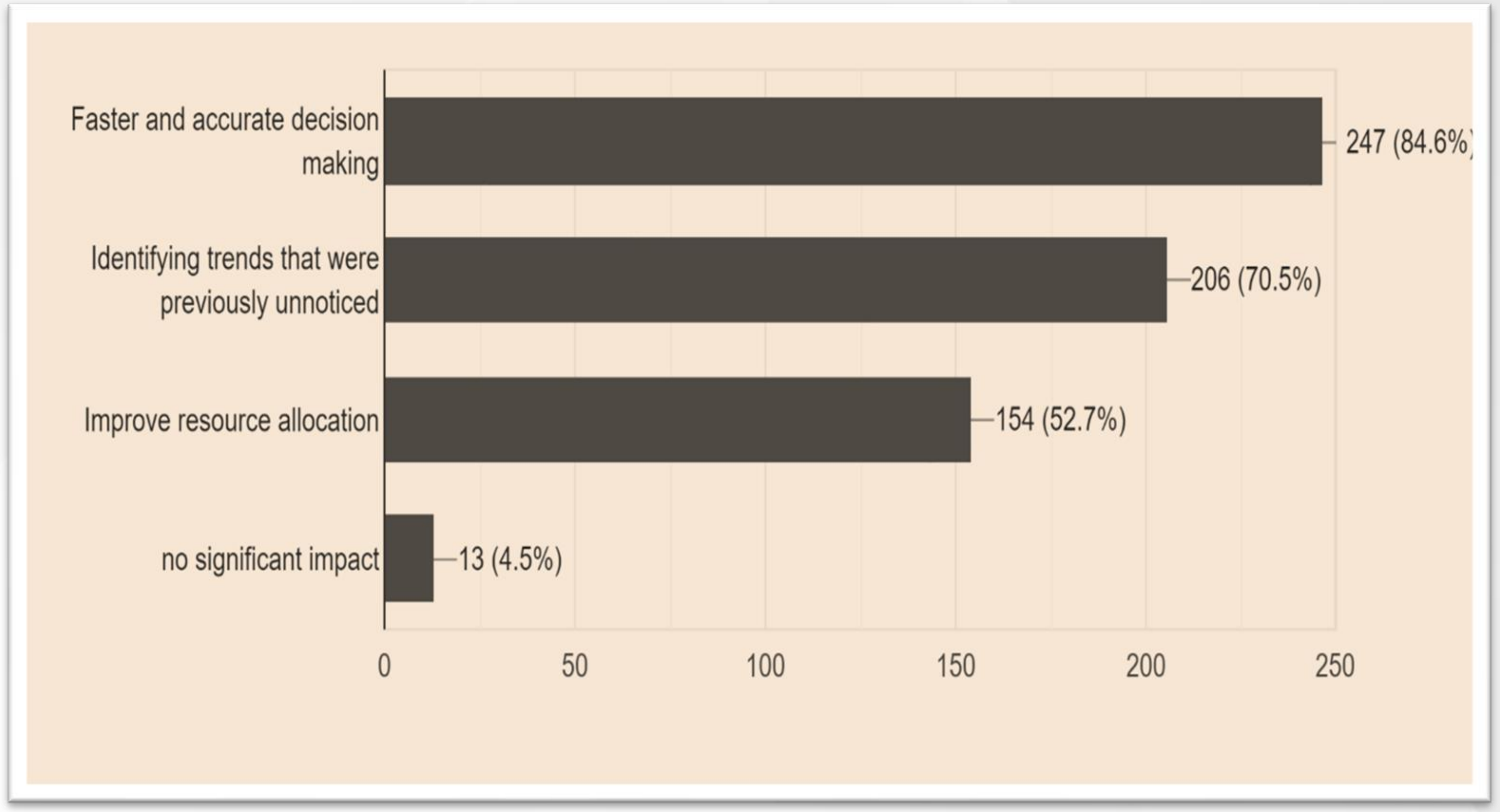




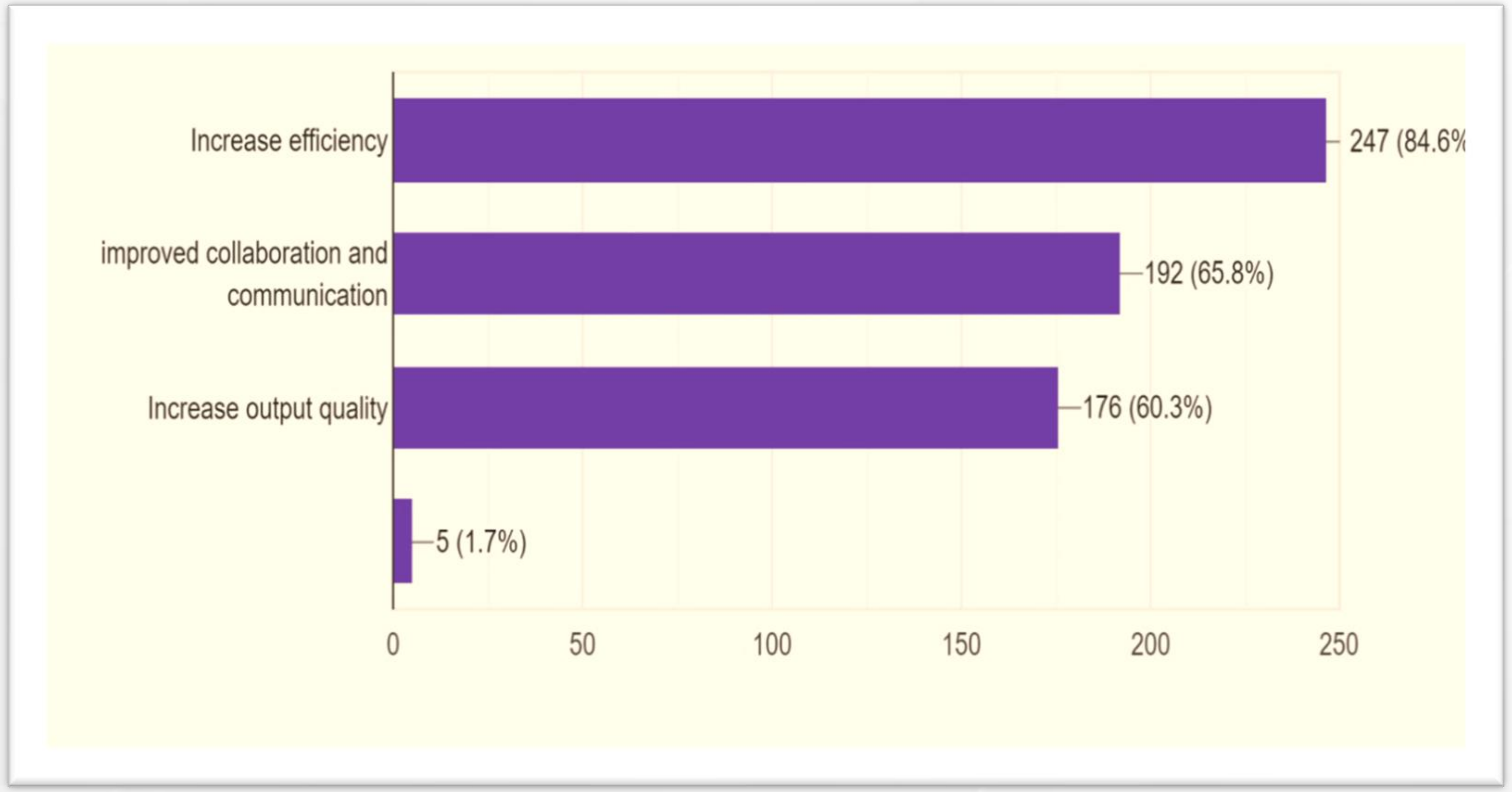
AI applications significantly impact on the efficiency in your organization



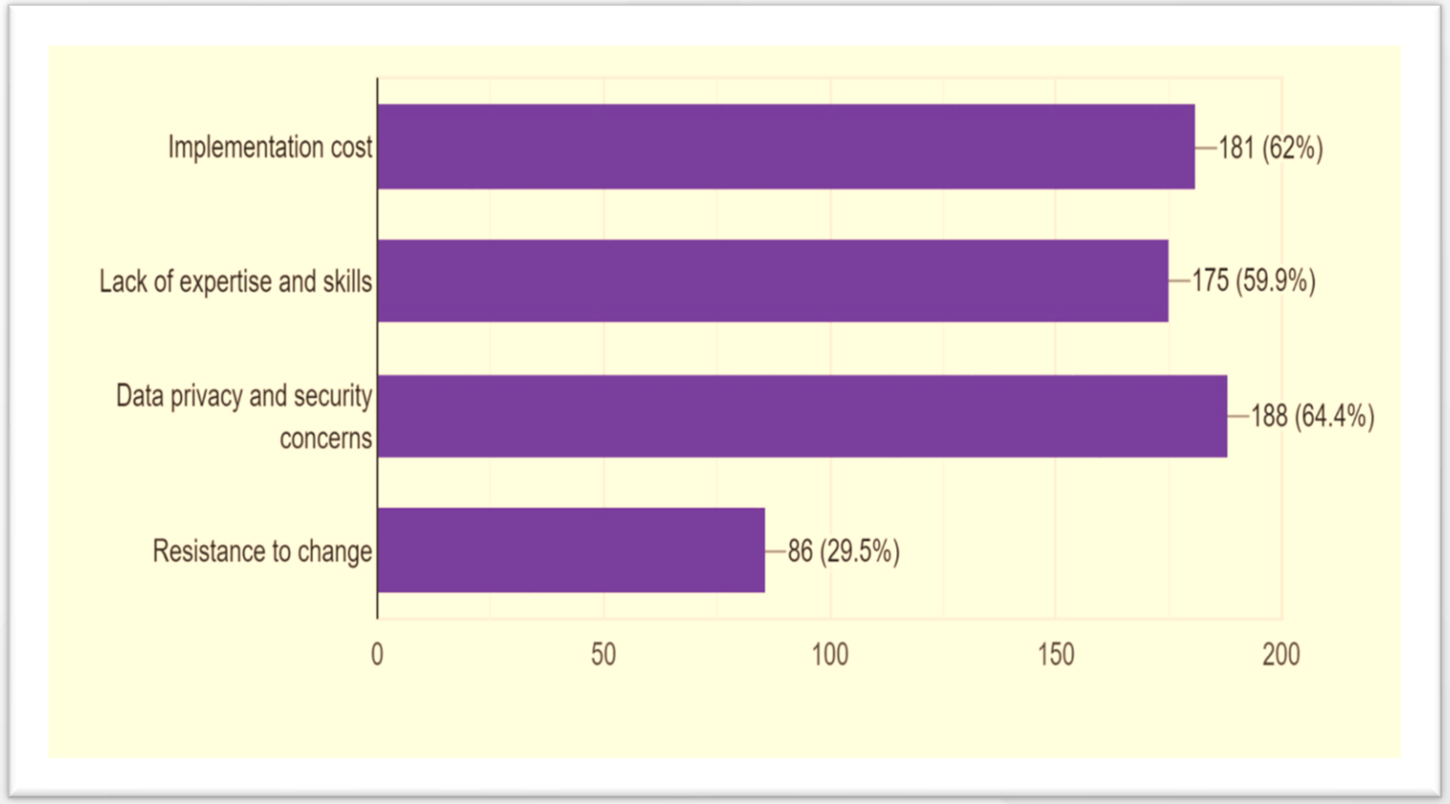
Data processing and analysis impacted productivity in the organization.



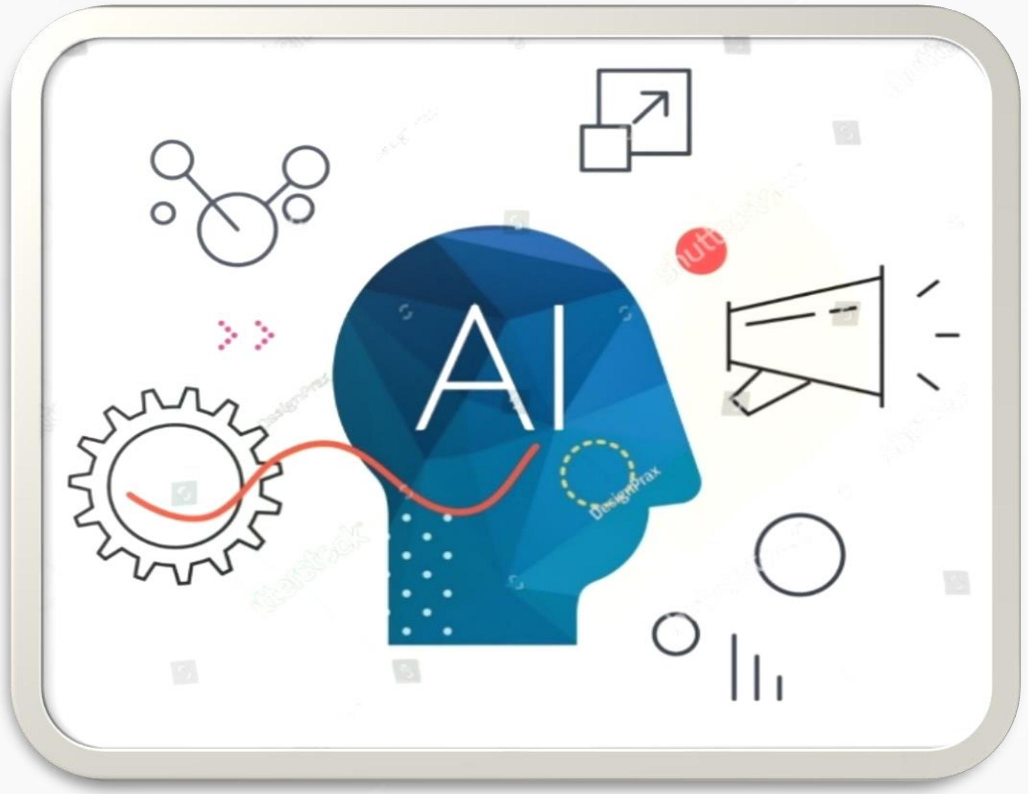
Workflow optimization impacted productivity in the organization.



The main challenges company faced while adopting AI .



Objective 1



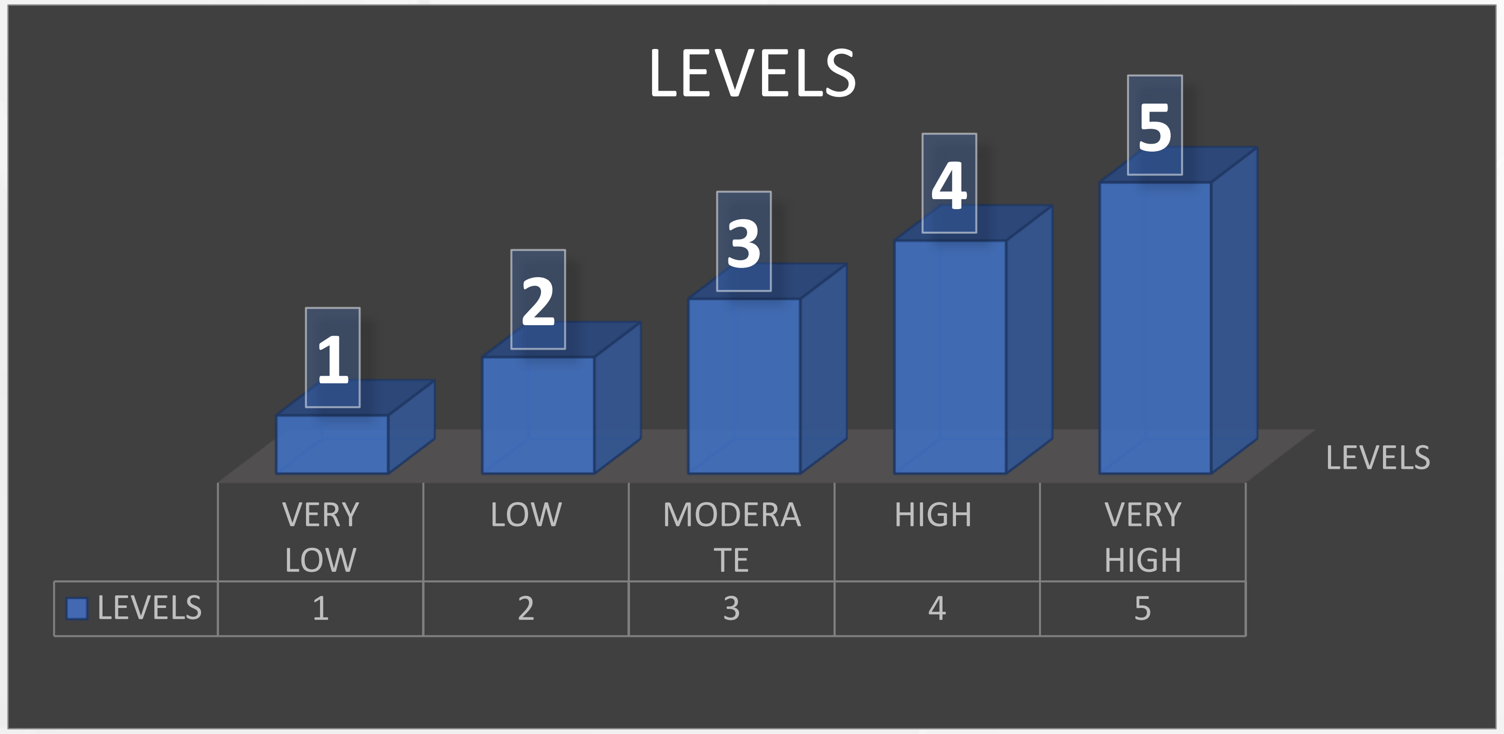
To study the impact of Artificial Intelligence on the Efficiency of the work in the Industries.

Variables used :-

✓ Outcome Variable (Y)– Efficiency of Work

In objective 1,the **Outcome Variable** as well as the **Predictor Variables** are fitted on **Ordinal Scale** having following levels –

1-Very low
2-Low
3-Moderate
4-High
5-Very high



✓ Predictor Variables

1) **X1-Workforce**
2) **X2-Capital investment**
3) **X3-Employee Productivity**
4) **X4-Workplace Environment**
5) **X5-Time Management**
6) **X6-Workload and Work life Balance**
7) **X7-Resource Availability**
8) **X8-Communication and Collaboration**
9) **X9-Task Automation**
10) **X10-Leadership and Management**

Analysis for 1st Objective

•(Full model Fitting)

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Y1 = 1]	-11.762	1.042	127.524	1	<.001	-13.803	-9.720
	[Y1 = 2]	-10.239	.911	126.242	1	<.001	-12.026	-8.453
	[Y1 = 3]	-6.582	.723	82.802	1	<.001	-7.999	-5.164
	[Y1 = 4]	-2.379	.569	17.507	1	<.001	-3.493	-1.264
Location	[X1=1]	-7.478	1.341	31.083	1	<.001	-10.106	-4.849
	[X1=2]	-2.994	.792	14.277	1	<.001	-4.547	-1.441
	[X1=3]	-3.167	.487	42.223	1	<.001	-4.122	-2.212
	[X1=4]	-1.782	.408	19.087	1	<.001	-2.581	-.982
	[X1=5]	0a	.	.	0	.	.	.
	[X2=1]	-2.087	1.446	2.083	1	.009	-4.921	.747
	[X2=2]	.245	.659	.138	1	.711	-1.048	1.537
	[X2=3]	-.383	.450	.724	1	.395	-1.264	.499
	[X2=4]	-.266	.407	.427	1	.513	-1.064	.532
	[X2=5]	0a	.	.	0	.	.	.
	[X3=1]	2.235	1.778	1.580	1	.209	-1.250	5.721
	[X3=2]	.082	.835	.010	1	.922	-1.554	1.718
	[X3=3]	-.871	.465	3.504	1	.006	-1.782	.041
	[X3=4]	-.207	.426	.235	1	.628	-1.042	.629
	[X3=5]	0a	.	.	0	.	.	.
	[X4=1]	.109	2.118	.003	1	.959	-4.041	4.259
	[X4=2]	-1.473	.865	2.902	1	.008	-3.168	.222
	[X4=3]	-.291	.484	.362	1	.547	-1.240	.657
	[X4=4]	-.891	.425	4.396	1	.036	-1.725	-.058
	[X4=5]	0a	.	.	0	.	.	.
	[X5=1]	-8.406	1.824	21.241	1	<.001	-11.981	-4.831
	[X5=2]	-2.962	.972	9.280	1	.002	-4.868	-1.056
	[X5=3]	.053	.455	.014	1	.906	-.839	.946
	[X5=4]	-.038	.395	.009	1	.923	-.812	.736
	[X5=5]	0a	.	.	0	.	.	.
	[X6=1]	-2.516	1.418	3.150	1	.006	-5.295	.263
	[X6=2]	-.844	.780	1.172	1	.279	-2.374	.685
	[X6=3]	-.410	.473	.752	1	.386	-1.337	.517
	[X6=4]	.432	.384	1.268	1	.260	-.320	1.184
	[X6=5]	0a	.	.	0	.	.	.
	[X7=1]	-.865	1.668	.269	1	.604	-4.134	2.404
	[X7=2]	-.322	.757	.180	1	.671	-1.805	1.162
	[X7=3]	.041	.489	.007	1	.934	-.917	.999
	[X7=4]	-.352	.440	.639	1	.424	-1.215	.511
	[X7=5]	0a	.	.	0	.	.	.
	[X8=1]	.072	1.883	.001	1	.970	-3.619	3.763
	[X8=2]	.045	.827	.003	1	.957	-1.577	1.666
	[X8=3]	-.715	.476	2.256	1	.133	-1.649	.218
	[X8=4]	-.310	.404	.589	1	.443	-1.101	.481
	[X8=5]	0a	.	.	0	.	.	.
	[X9=1]	-18.995	.000	.	1	.	-18.995	-18.995
	[X9=2]	-3.384	.884	14.654	1	<.001	-5.116	-1.651
	[X9=3]	-1.465	.449	10.633	1	.001	-2.345	-.584
	[X9=4]	-.449	.368	1.492	1	.222	-1.170	.272
	[X9=5]	0a	.	.	0	.	.	.
	[X10=1]	.414	1.122	.136	1	.712	-1.786	2.614
	[X10=2]	-.068	.637	.011	1	.915	-1.316	1.181
	[X10=3]	-.334	.492	.462	1	.497	-1.298	.630
	[X10=4]	.348	.416	.700	1	.403	-.467	1.163
	[X10=5]	0a	.	.	0	.	.	.

Link function: Logit

(Out of the five levels one level is taken as the reference category ,here the level 5 ,i.e. very high is taken as the reference category)

From full model we get 7 variables significant i.e. X1, X2, X3, X4, X5, X6, X9

X1=Workforce

X2=Capital Investment

X3= Employee Predictivity

X4=Workplace Environment

X5=Time Management

X6=Workload and Work Life Balance

X9=Task automation

- ## Proportional odds assumption-

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	421.901			
General	338.753[b]	83.148[c]	120	.996

The above assumption represents parallel line logistic regression model ,i.e. all Slopes are equal based on cumulative distribution probability of response level. We use score test to test proportional odds assumption, i.e. we test whether or not ordinal restrictions are valid.

We test,

H0: $\beta_{1l} = \beta_1 \quad i.e. \quad \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = \beta_1 \quad \forall \; l = 1, 2, ., k-1$

H1: $\beta_{1l} \neq \beta_1$ for atleast one l

Where β_1 depicts common slope parameter.

We Hypothesis that there is common slope parameter

Interpretation-

P-value = 0.996 > 0.05

Therefore, we do not reject Null hypothesis and conclude that there is common slope parameter for each outcome variable.

- ## To check Multicollinearity-

	GVIF	Df	GVIF^(1/(2*Df))
Y1	4.241951	4	1.197969
X1	1.867255	1	1.366475
X2	2.799336	4	1.137318
X3	3.012751	4	1.147811
X4	2.437153	1	1.561138
X5	1.877715	1	1.370297
X6	2.368867	1	1.539112
X7	3.813546	4	1.182132
X8	3.871013	4	1.184345
X9	1.786982	1	1.336781
X10	3.817951	4	1.182303

GVIF (Generalized Variance Inflation Factor) is a measure used to detect multicollinearity in regression models, particularly when dealing with categorical predictors. It extends the concept of **VIF** (Variance Inflation Factor) to handle the complexity of models with categorical variables.GVIF can be adjusted for the number of degrees of freedom (df) associated with the predictor, resulting in a measure called **GVIF^(1/(2*df))**. This adjustment helps interpret GVIF in models with categorical variables having multiple levels.

GVIF^(1/(2*df)) close to 1: Indicates little to no multicollinearity.

GVIF^(1/(2*df)) greater than 1: Indicates the presence of multicollinearity, with higher values suggesting more severe multicollinearity.

Interpretation:

In the above table all the adjusted GVIF values are close to 1 which indicates relatively little to no collinearity which is acceptable.

❖ Reduced model fitting)

➢ PARAMETER ESTIMATES-

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Y1 = 1]	-11.424	.993	132.455	1	<.001	-13.370	-9.479
	[Y1 = 2]	-9.921	.853	135.356	1	<.001	-11.592	-8.249
	[Y1 = 3]	-6.373	.665	91.751	1	<.001	-7.677	-5.069
	[Y1 = 4]	-2.292	.513	19.942	1	<.001	-3.298	-1.286
Location	[X1=1]	-7.036	1.298	29.394	1	<.001	-9.579	-4.492
	[X1=2]	-3.130	.719	18.926	1	<.001	-4.540	-1.720
	[X1=3]	-3.177	.473	45.113	1	<.001	-4.104	-2.250
	[X1=4]	-1.740	.389	19.992	1	<.001	-2.502	-.977
	[X1=5]	0[a]	.	.	0	.	.	.
	[X4=1]	.111	1.944	.003	1	.954	-3.699	3.922
	[X4=2]	-1.703	.751	5.149	1	.003	-3.174	-.232
	[X4=3]	-.524	.428	1.503	1	.220	-1.362	.314
	[X4=4]	-.812	.388	4.390	1	.036	-1.571	-.052
	[X4=5]	0[a]	.	.	0	.	.	.
	[X5=1]	-7.694	1.541	24.932	1	<.001	-10.715	-4.674
	[X5=2]	-2.691	.916	8.626	1	.003	-4.486	-.895
	[X5=3]	.127	.436	.084	1	.771	-.729	.982
	[X5=4]	-.077	.387	.040	1	.841	-.835	.680
	[X5=5]	0[a]	.	.	0	.	.	.
	[X6=1]	-2.180	1.230	3.144	1	.006	-4.591	.230
	[X6=2]	-.688	.726	.898	1	.343	-2.111	.735
	[X6=3]	-.569	.435	1.710	1	.191	-1.423	.284
	[X6=4]	.366	.370	.979	1	.322	-.359	1.090
	[X6=5]	0[a]	.	.	0	.	.	.
	[X9=1]	-19.344	.000	.	1	.	-19.344	-19.344
	[X9=2]	-3.397	.814	17.412	1	<.001	-4.993	-1.802
	[X9=3]	-1.544	.436	12.561	1	<.001	-2.398	-.690
	[X9=4]	-.490	.362	1.830	1	.176	-1.200	.220
	[X9=5]	0[a]	.	.	0	.	.	.
	[X2=1]	-2.084	1.377	2.291	1	.130	-4.782	.614
	[X2=2]	.083	.628	.018	1	.894	-1.147	1.314
	[X2=3]	-.511	.412	1.534	1	<.001	-1.319	.297
	[X2=4]	-.400	.380	1.109	1	.292	-1.144	.344
	[X2=5]	0[a]	.	.	0	.	.	.
	[X3=1]	2.176	1.825	1.422	1	.233	-1.400	5.752
	[X3=2]	.088	.811	.012	1	.913	-1.502	1.679
	[X3=3]	-.988	.456	4.702	1	<.001	-1.882	-.095
	[X3=4]	-.280	.408	.470	1	.493	-1.080	.520
	[X3=5]	0[a]	.	.	0	.	.	.
Link function: Logit.								

$$\Pi_1 = \frac{e^{logit\ F1}}{1 + e^{logitF1}}$$

where,

LogitF1=-11.424-7.036*Workforce(1)-3.130*Workforce(2) -3.177*Workforce(3)-1.740*Workforce(4) -2.084*Capital investment(1)+0.083*Capital investment(2)-0.511*Capital investment(3)-0.400*Capital investment(4)+2.176*Employee productivity(1)+0.088* Employee productivity(2)-0.988* Employee productivity(3)-0.280* Employee productivity(4)+0.111*Workplace environment(1)-1.703*Workplace environment(2)-0.524*Workplace environment(3)-0.812*Workplace environment(4)-7.694*Time management(1)-2.691*Time management(2)0.127*Time management(3)-0.077*Time management(4)-2.180*Workload and Work life balance(1)-0.688*Workload and Work life balance(2)-0.569*Workload and Work life balance(3)+0.366*Workload and Work life balance(4)-19.344*Task automation(1)-3.397*Task automation(2)-1.544*Task automation(3)-4.90*Task automation(4)

$$\Pi_2 = \frac{e^{logit\ F2}}{1 + e^{logitF2}}$$

where,

LogitF2 =-9.921-7.036*Workforce(1)-3.130*Workforce(2) -3.177*Workforce(3)-1.740*Workforce(4) -2.084*Capital investment(1)+0.083*Capital investment(2)-0.511*Capital investment(3)-0.400*Capital investment(4)+2.176*Employee productivity(1)+0.088* Employee productivity(2)-0.988* Employee productivity(3)-0.280* Employee productivity(4)+0.111*Workplace environment(1)-1.703*Workplace environment(2)-0.524*Workplace environment(3)-0.812*Workplace environment(4)-7.694*Time management(1)-2.691*Time management(2)0.127*Time management(3)-0.077*Time management(4)-2.180*Workload and Work life balance(1)-0.688*Workload and Work life balance(2)-0.569*Workload and Work life balance(3)+0.366*Workload and Work life balance(4)-19.344*Task automation(1)-3.397*Task automation(2)-1.544*Task automation(3)-4.90*Task automation(4)

$$\Pi_3 = \frac{e^{logit\ F3}}{1 + e^{logitF3}}$$

where,

LogitF3 =-6.373-7.036*Workforce(1)-3.130*Workforce(2) -3.177*Workforce(3)-1.740*Workforce(4) -2.084*Capital investment(1)+0.083*Capital investment(2)-0.511*Capital investment(3)-0.400*Capital investment(4)+2.176*Employee productivity(1)+0.088* Employee productivity(2)-0.988* Employee productivity(3)-0.280* Employee productivity(4)+0.111*Workplace environment(1)-1.703*Workplace environment(2)-0.524*Workplace environment(3)-0.812*Workplace environment(4)-7.694*Time management(1)-2.691*Time management(2)0.127*Time management(3)-0.077*Time management(4)-2.180*Workload and Work life balance(1)-0.688*Workload and Work life balance(2)-0.569*Workload and Work life balance(3)+0.366*Workload and Work life balance(4)-19.344*Task automation(1)-3.397*Task automation(2)-1.544*Task automation(3)-4.90*Task automation(4)

$$\Pi_4 = \frac{e^{logit\ F4}}{1 + e^{logitF4}}$$

where,

LogitF4 =-2.292-7.036*Workforce(1)-3.130*Workforce(2) -3.177*Workforce(3)-1.740*Workforce(4) -2.084*Capital investment(1)+0.083*Capital investment(2)-0.511*Capital investment(3)-0.400*Capital investment(4)+2.176*Employee productivity(1)+0.088* Employee productivity(2)-0.988* Employee productivity(3)-0.280* Employee productivity(4)+0.111*Workplace environment(1)-1.703*Workplace environment(2)-0.524*Workplace environment(3)-0.812*Workplace environment(4)-7.694*Time management(1)-2.691*Time management(2)0.127*Time management(3)-0.077*Time management(4)-2.180*Workload and Work life balance(1)-0.688*Workload and Work life balance(2)-0.569*Workload and Work life balance(3)+0.366*Workload and Work life balance(4)-19.344*Task automation(1)-3.397*Task automation(2)-1.544*Task automation(3)-4.90*Task automation(4)

Proportional odds assumption-

Test of Parallel Lines				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	417.499			
General	351.845[b]	65.654[c]	84	.931

We test,
H0: $\beta_{1l} = \beta_1$ $i.e.$ $\beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = \beta_1 \quad \forall \quad l = 1, 2, \dots, k-1$
H1: $\beta_{1l} \neq \beta_1$ for atleast one l
Where β_1 depicts common slope parameter.
We Hypothesis that there is common slope parameter
Interpretation-P-value = 0.931 > 0.05
Therefore, we do not reject Null hypothesis and conclude that there is common slope parameter for each outcome variable.

- Model fitting information and Goodness-of-fit-test-

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	699.945			
Final	417.499	282.446	28	<.001
Link function: Logit.				

Hypothesis:
H_0: The predictor variables do not significantly improve the fit of the model compared to the null model (without predictors)
H_1: At least one of the predictor variables significantly improves the fit of the model
Interpretation: P-value = 0.00 < 0.05 .Therefore, we reject Null hypothesis and say that the predictor variables significantly improves the fit of the model.

❖ Likelihood Ratio test-
Model 1: Y1 ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X 10
Model 2: Y1 ~ X1 + X2 + X3 + X4 + X5 + X6 + X9

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	**44**	**-212.34**			
2	**32**	**-216.90**	**12**	**9.1206**	**0.6926**

Hypothesis :H_0:Reduced model is as good as full model
H_1 :Reduced model is not as good as full model
Interpretation: Here the P-value =0.6926>0.05.
Hence, we do not reject the null hypothesis and conclude that Reduced model is as good as Full model.

- Goodness of fit test-

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	680.588	884	1.000
Deviance	404.288	884	1.000
Link function: Logit.			

This includes tests like Pearson and Deviance, which assess how well the model fits the data. If p-value > 0.05 it indicates a good fit.
Hypothesis:H_0: The model fits the data well.
H_1: The model does not fit data well
Interpretation –Here P-value – 1.000 >0.05.Therefore, we do not reject Null hypothesis and say that the model fits the data well.

To test Accuracy

Full model accuracy

Y1 * Predicted Response Category Crosstabulation

Count							
		Predicted Response Category					Total
		1	2	3	4	5	
Y1	1	3	3	2	0	0	8
	2	1	0	0	0	0	1
	3	4	6	30	9	0	49
	4	0	0	21	110	30	161
	5	0	0	0	20	53	73
Total		8	9	53	139	83	292

The accuracy of the full model is 67.12%.

Reduced model accuracy

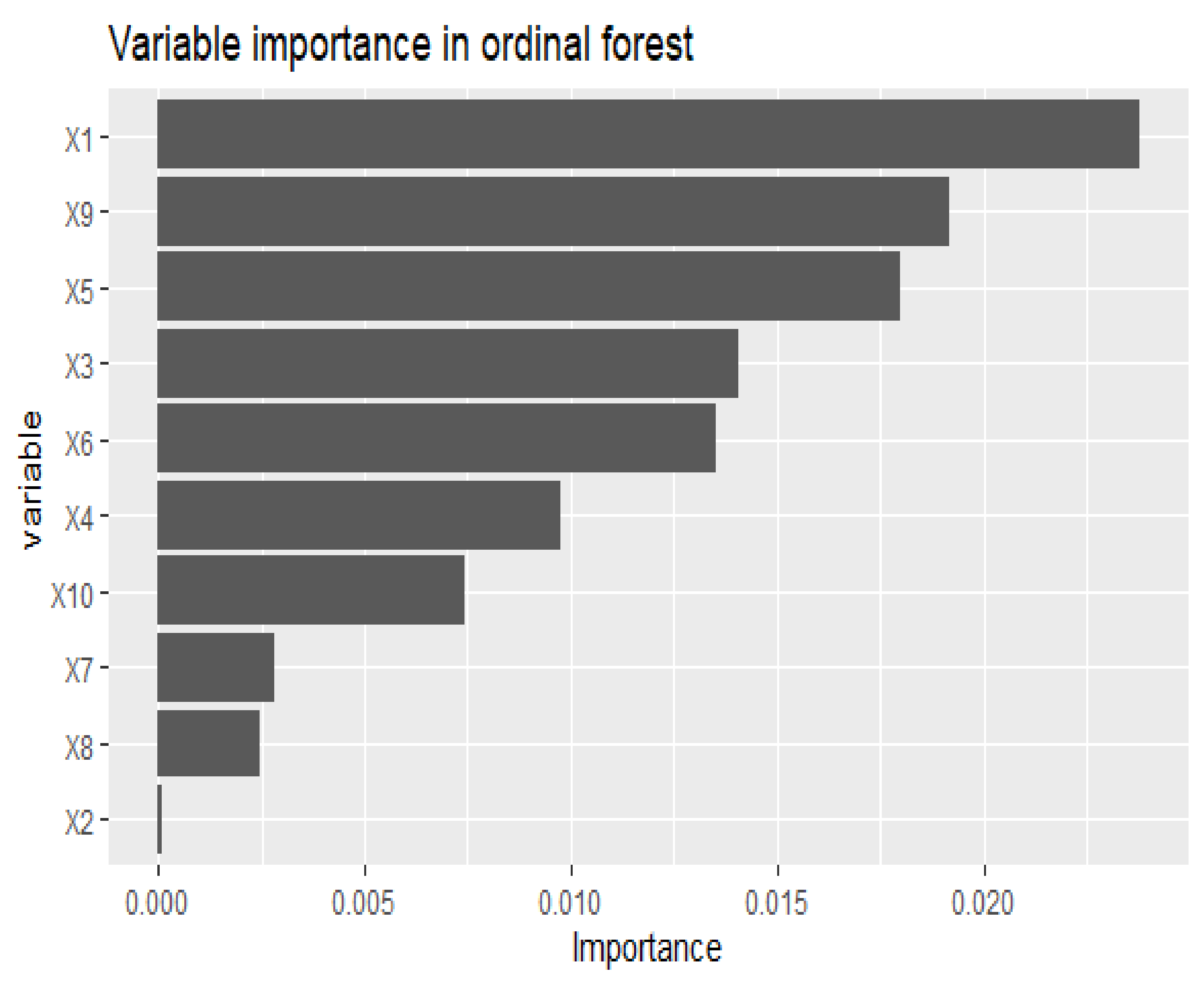
Y1 * Predicted Response Category Crosstabulation

Count							
		Predicted Response Category					Total
		1	2	3	4	5	
Y1	1	2	3	2	0	0	7
	2	2	0	0	0	0	2
	3	4	6	30	3	0	43
	4	0	0	19	117	31	167
	5	0	0	2	19	52	73
Total		8	9	53	139	83	292

The accuracy of the Reduced model is 68.84%.

➢ Analysis done by Ordinal Forest technique

	variable Importance
X1	**X1 2.378499e-02**
X2	**X2 9.950023e-05**
X3	**X3 1.405385e-02**
X4	**X4 9.771447e-03**
X5	**X5 1.796626e-02**
X6	**X6 1.353314e-02**
X7	**X7 2.828807e-03**
X8	**X8 2.458726e-03**
X9	**X9 1.916054e-02**
X10	**X10 7.419457e-03**



Interpretation-From the table of importance and graph we can see that the variable X1,X9,X5 has the most importance compared to rest of the variables.

To test Accuracy

Training accuracy of the fitted model

pred_labels	1	2	3	4	5
1	3	0	0	0	0
2	3	6	0	0	0
3	0	0	39	3	0
4	0	0	1	84	0
5	0	0	0	11	62

Accuracy=0.9150943

Testing accuracy of the fitted model

pred_labels	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	2	0	7	2	0
4	0	1	5	26	2
5	0	0	1	13	19

Accuracy=0.675

The accuracy of the training model is 91.50%

The accuracy of the testing model is 67.50%

Method	Model	Accuracy
Ordinal Logistic Regression	Full Model Reduced Model	67.12% 68.84%
Ordinal Forest	Full Model (Trained Data) Full Model(Testing data)	91.50% 67.50%

Conclusion For Objective 1-

So from the best fitted model ,i.e.reduced model we get 7 variables significant,i.e. X1,X2,X3,X4,X5,X6,X9 which are workforce , capital investment ,Employee productivity, Workplace environment,time management, workload and work life balance, task automation .
From this data analysis, we get know that impact of AI on Efficiency is level 4 ,i.e. HIGH.
Also to increase the accuracy of the model we used machine learning technique ,i.e ordinal forest technique which gives us the accuracy 91.50% on training data and 67.50% on testing data which gives better accuracy ,hence we predict our result from model fitted by ordinal forest technique which gives us the result level 4,i.e HIGH by taking the mode of the predictions.Hence ,the impact of AI on Efficiency of work in industries is HIGH.

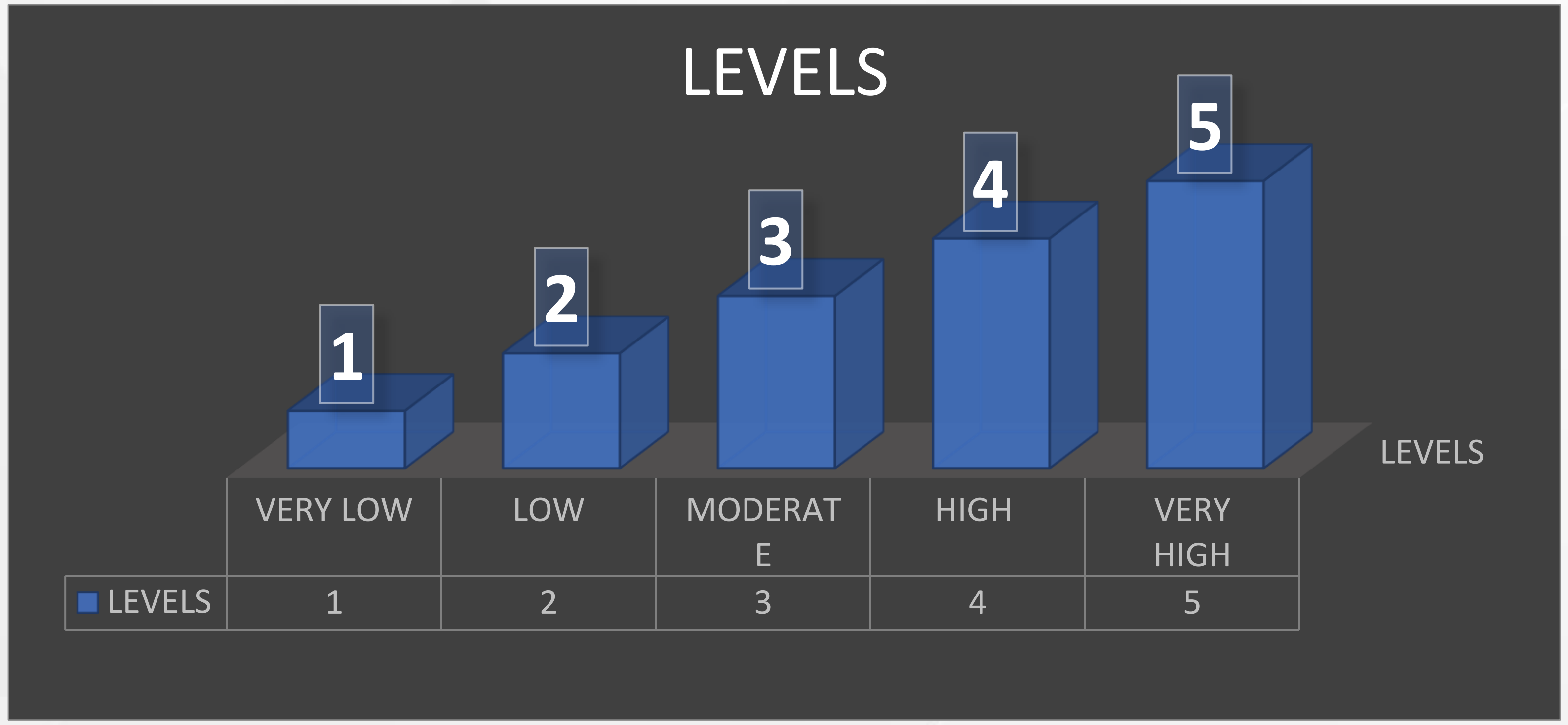
Objective 2

To analyze the impact of Artificial Intelligence on the Expenditure in the Industries.

<u>Variables used</u> :-

✓<u>Outcome Variable</u> (Y)– <u>Expenditure</u>

In objective 2,the **<u>Outcome Variable</u>** as well as the **<u>Predictor Variables</u>** are fitted on **<u>Ordinal Scale</u>** having following levels –
1-Very low
2-Low
3-Moderate
4-High
5-Very high



✓<u>Predictor Variables</u>

1) XI- Resource Management
2) X2- Strategic sourcing
3) X3- Labour cost Management
4) X4-Risk Management
5) X5- Customer Relationship Management
6) X6-Supply Chain Management
7) X7-Operational Efficiency
8) X8-Energy Efficiency
9) X9-Product Design

Analysis of 2nd objective

(To increase the accuracy of the model we divide data into 80% and 20%)

- (Full model fitting)-
- For training data:

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Y2 = 1]	-10.473	1.049	99.743	1	<.001	-12.528	-8.418
	[Y2 = 2]	-8.142	.834	95.372	1	<.001	-9.776	-6.508
	[Y2 = 3]	-4.574	.719	40.425	1	<.001	-5.984	-3.164
	[Y2 = 4]	-2.201	.649	11.508	1	<.001	-3.473	-.929
Location	[X1=1]	-10.831	3.182	11.584	1	<.001	-17.068	-4.594
	[X1=2]	-2.146	.811	6.996	1	.008	-3.736	-.556
	[X1=3]	-1.189	.590	4.064	1	.044	-2.345	-.033
	[X1=4]	-.756	.527	2.060	1	.151	-1.788	.276
	[X1=5]	0[a]	.	.	0	.	.	.
	[X2=1]	8.250	3.050	7.315	1	.007	2.271	14.228
	[X2=2]	1.180	.817	2.086	1	.149	-.421	2.781
	[X2=3]	.510	.518	.971	1	.324	-.505	1.525
	[X2=4]	.412	.444	.864	1	.353	-.457	1.282
	[X2=5]	0[a]	.	.	0	.	.	.
	[X3=1]	-6.152	1.563	15.500	1	<.001	-9.214	-3.089
	[X3=2]	-1.318	.684	3.711	1	.054	-2.658	.023
	[X3=3]	-.590	.519	1.294	1	.255	-1.607	.427
	[X3=4]	.310	.461	.453	1	.501	-.594	1.215
	[X3=5]	0[a]	.	.	0	.	.	.
	[X4=1]	1.724	1.121	2.365	1	.124	-.473	3.921
	[X4=2]	-.759	.785	.933	1	.334	-2.297	.780
	[X4=3]	-.775	.488	2.522	1	.112	-1.732	.181
	[X4=4]	.097	.466	.043	1	.835	-.817	1.011
	[X4=5]	0[a]	.	.	0	.	.	.
	[X5=1]	0[a]	.	.	0	.	.	.
	[X5=2]	-2.401	.757	10.048	1	.002	-3.885	-.916
	[X5=3]	-1.091	.534	4.172	1	.041	-2.138	-.044
	[X5=4]	-.606	.464	1.711	1	.191	-1.515	.302
	[X5=5]	0[a]	.	.	0	.	.	.
	[X6=1]	.363	1.594	.052	1	.820	-2.761	3.487
	[X6=2]	-3.598	.973	13.672	1	<.001	-5.505	-1.691
	[X6=3]	-1.007	.499	4.075	1	.044	-1.984	-.029
	[X6=4]	-1.224	.438	7.804	1	.005	-2.082	-.365
	[X6=5]	0[a]	.	.	0	.	.	.
	[X7=1]	1.257	1.367	.845	1	.358	-1.423	3.936
	[X7=2]	1.160	1.078	1.158	1	.282	-.952	3.272
	[X7=3]	.715	.528	1.833	1	.176	-.320	1.750
	[X7=4]	-.047	.474	.010	1	.921	-.976	.882
	[X7=5]	0[a]	.	.	0	.	.	.
	[X8=1]	-2.979	1.600	3.465	1	.063	-6.115	.158
	[X8=2]	-1.992	1.193	2.786	1	.095	-4.330	.347
	[X8=3]	-2.376	.499	22.697	1	<.001	-3.354	-1.399
	[X8=4]	-1.216	.435	7.805	1	.005	-2.069	-.363
	[X8=5]	0[a]	.	.	0	.	.	.
	[X9=1]	-4.328	1.473	8.628	1	.003	-7.216	-1.440
	[X9=2]	1.753	.830	4.459	1	.035	.126	3.380
	[X9=3]	-.030	.536	.003	1	.955	-1.080	1.020
	[X9=4]	.197	.411	.231	1	.631	-.607	1.002
	[X9=5]	0[a]	.	.	0	.	.	.

Link function: Logit

(Out of the five levels one level is taken as the reference category ,here the level 5 ,i.e. very high is taken as the reference category)

From full model we get 7 variables significant i.e. X1, X2, X3, X5, X6, X8,X9

X1= Resource Management

X2=Strategic sourcing

X3= Labour cost Management

X5=Customer Relationship Management

X6=Supply Chain Management

X8=Energy Efficiency

X9=Product Design

• Checking proportional odds assumption of full model:

Hypothesis- H_0: $\beta 1l = \beta 1$ $\forall$ $l=1, 2, 3, 4$

H_1: $\beta 1l \neq \beta 1$ for atleast one l

Test of Parallel Lines[a]

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	384.121			
General	297.635[b]	86.486[c]	105	.906

We Hypothesis that there is common slope parameter for each of the cumulative logit regression equation instead of 4 distinct slopes.

Interpretation-P-value = 0.906 > 0.05.Therefore, we do not reject Null hypothesis and conclude that there is common slope parameter for each outcome variable.

➢ To check Multicollinearity:

GVIF	Df	GVIF^(1/(2*Df))
Y2 2.275176	4	1.108222
X1 2.571435	1	1.603569
X2 2.023901	1	1.422639
X3 2.221589	1	1.490500
X4 1.803771	1	1.343046
X5 2.037464	1	1.427398
X6 1.674401	1	1.293987
X7 1.848889	1	1.359738
X8 1.932334	1	1.390084
X9 1.921527	1	1.386191

GVIF^(1/(2*df)) close to 1: Indicates little to no multicollinearity. **GVIF^(1/(2*df)) greater than 1:** Indicates the presence of multicollinearity, with higher values suggesting more severe multicollinearity.

Interpretation: In the above table all the adjusted GVIF values are close to 1 which indicates relatively little to no collinearity which is acceptable.

❖(Reduced model fitting):

(For reduced model we only consider significant variables from the full model)

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Y2 = 1]	-10.281	1.018	101.892	1	<.001	-12.277	-8.285
	[Y2 = 2]	-7.890	.781	102.158	1	<.001	-9.420	-6.360
	[Y2 = 3]	-4.483	.668	44.982	1	<.001	-5.793	-3.173
	[Y2 = 4]	-2.235	.600	13.898	1	<.001	-3.410	-1.060
Location	[X1=1]	-10.300	3.148	10.705	1	.001	-16.470	-4.130
	[X1=2]	-2.179	.768	8.047	1	.005	-3.685	-.674
	[X1=3]	-1.326	.559	5.631	1	.018	-2.421	-.231
	[X1=4]	-.793	.498	2.536	1	.111	-1.769	.183
	[X1=5]	0[a]	.	.	0	.	.	.
	[X2=1]	7.667	2.882	7.080	1	.008	2.020	13.315
	[X2=2]	.710	.774	.842	1	.359	-.807	2.227
	[X2=3]	.451	.503	.802	1	.370	-.536	1.437
	[X2=4]	.279	.427	.427	1	.513	-.558	1.115
	[X2=5]	0[a]	.	.	0	.	.	.
	[X3=1]	-5.413	1.466	13.624	1	<.001	-8.287	-2.539
	[X3=2]	-1.365	.634	4.635	1	.031	-2.609	-.122
	[X3=3]	-.462	.504	.840	1	.359	-1.450	.526
	[X3=4]	.231	.445	.269	1	.604	-.642	1.104
	[X3=5]	0[a]	.	.	0	.	.	.
	[X5=1]	0[a]	.	.	0	.	.	.
	[X5=2]	-1.835	.665	7.606	1	.006	-3.139	-.531
	[X5=3]	-.969	.509	3.630	1	.057	-1.966	.028
	[X5=4]	-.497	.446	1.243	1	.265	-1.372	.377
	[X5=5]	0[a]	.	.	0	.	.	.
	[X6=1]	.050	1.534	.001	1	.974	-2.956	3.057
	[X6=2]	-3.623	.901	16.173	1	<.001	-5.389	-1.857
	[X6=3]	-1.000	.472	4.489	1	.034	-1.924	-.075
	[X6=4]	-1.103	.429	6.601	1	.010	-1.945	-.262
	[X6=5]	0[a]	.	.	0	.	.	.
	[X8=1]	-2.391	1.382	2.995	1	.083	-5.099	.317
	[X8=2]	-1.796	1.138	2.489	1	.115	-4.027	.435
	[X8=3]	-2.441	.482	25.694	1	<.001	-3.385	-1.497
	[X8=4]	-1.183	.408	8.396	1	.004	-1.984	-.383
	[X8=5]	0[a]	.	.	0	.	.	.
	[X9=1]	-3.328	1.389	5.737	1	.017	-6.051	-.605
	[X9=2]	1.668	.780	4.576	1	.032	.140	3.196
	[X9=3]	.127	.511	.062	1	.804	-.874	1.128
	[X9=4]	.145	.401	.131	1	.718	-.640	.930
	[X9=5]	0[a]	.	.	0	.	.	.

Ordinal logistic regression model based on parameter estimates by the reduced model.

$$\Pi_1 = \frac{e^{logit\ F1}}{1+e^{logitF1}}$$

Where , **logit F1**= -10.281-10.300*Resource management1- 2.179* Resource management 2- 1.326*Resource management3-0.793* Resource management4+7.667*Strategic sourcing1+0.710 *Strategic sourcing2+0.451 *Strategic sourcing3+0.279 *Strategic sourcing4-5.413*Labor cost management1-1.365* Labor cost management2-0.462* Labor cost management3+0.231*Labor cost management4-1.835*Customer relationship management2-0.969* Customer relationship management3-0.497 *Customer relationship management4+0.50*Supply chain management1-3.623* Supply chain management2-1.000* Supply chain management3-1.103 *Supply chain management4-2.391*Energy efficiency1-1.796* Energy efficiency2-2.441* Energy efficiency3-1.183 *Energy efficiency4-3.328* Product design1+1.668* Product design2+0.127* Product design3+0.145 *Product design4

$$\Pi_2 = \frac{e^{logit\ F2}}{1+e^{logitF2}}$$

Where , **logit F2**= -7.890-10.300*Resource management1- 2.179* Resource management 2- 1.326*Resource management3-0.793* Resource management4+7.667*Strategic sourcing1+0.710 *Strategic sourcing2+0.451 *Strategic sourcing3+0.279 *Strategic sourcing4-5.413*Labor cost management1-1.365* Labor cost management2-0.462* Labor cost management3+0.231*Labor cost management4-1.835*Customer relationship management2-0.969* Customer relationship management3-0.497 *Customer relationship management4+0.50*Supply chain management1-3.623* Supply chain management2-1.000* Supply chain management3-1.103 *Supply chain management4-2.391*Energy efficiency1-1.796* Energy efficiency2-2.441* Energy efficiency3-1.183 *Energy efficiency4-3.328* Product design1+1.668* Product design2+0.127* Product design3+0.145 *Product design4

$$\Pi_3 = \frac{e^{logit\ F3}}{1+e^{logitF3}}$$

Where , **logit F3**= -4.483-10.300*Resource management1- 2.179* Resource management 2- 1.326*Resource management3-0.793* Resource management4+7.667*Strategic sourcing1+0.710 *Strategic sourcing2+0.451 *Strategic sourcing3+0.279 *Strategic sourcing4-5.413*Labor cost management1-1.365* Labor cost management2-0.462* Labor cost management3+0.231*Labor cost management4-1.835*Customer relationship management2-0.969* Customer relationship management3-0.497 *Customer relationship management4+0.50*Supply chain management1-3.623* Supply chain management2-1.000* Supply chain management3-1.103 *Supply chain management4-2.391*Energy efficiency1-1.796* Energy efficiency2-2.441* Energy efficiency3-1.183 *Energy efficiency4-3.328* Product design1+1.668* Product design2+0.127* Product design3+0.145 *Product design4

$$\Pi_4 = \frac{e^{logit\ F4}}{1+e^{logitF4}}$$

Where , **logit F4**= -2.235-10.300*Resource management1- 2.179* Resource management 2- 1.326*Resource management3-0.793* Resource management4+7.667*Strategic sourcing1+0.710 *Strategic sourcing2+0.451 *Strategic sourcing3+0.279 *Strategic sourcing4-5.413*Labor cost management1-1.365* Labor cost management2-0.462* Labor cost management3+0.231*Labor cost management4-1.835*Customer relationship management2-0.969* Customer relationship management3-0.497 *Customer relationship management4+0.50*Supply chain management1-3.623* Supply chain management2-1.000* Supply chain management3-1.103 *Supply chain management4-2.391*Energy efficiency1-1.796* Energy efficiency2-2.441* Energy efficiency3-1.183 *Energy efficiency4-3.328* Product design1+1.668* Product design2+0.127* Product design3+0.145 *Product design4

- **Checking proportional odds assumption of reduced model:**

Hypothesis-H_0: $\beta_{1l} = \beta_1 \quad \forall \quad l = 1, 2, 3, 4$

H_1: $\beta_{1l} \neq \beta_1$ for atleast one l

We Hypothesis that there is common slope parameter for each of the cumulative logit regression equation instead of 4 distinct slopes.

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	382.511			
General	302.398[b]	80.113[c]	81	.507

Interpretation-P-value = 0.507 > 0.05.Therefore, we do not reject Null hypothesis and conclude that there is common slope parameter for each outcome variable.

- **Model Fitting Information**

Hypothesis:

H_0: The predictor variables do not significantly improve the fit of the model compared to the null model (without predictors)

H_1: At least one of the predictor variables significantly improves the fit of the model

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
InterceptOnly	590.377			
Final	382.511	207.866	27	<.001

Interpretation: P-value = 0.001 < 0.05.Therefore, we reject Null hypothesis and say the predictor variables significantly improves the fit of the model.

- **Likelihood ratio test :-**

Hypothesis-H0:Reduced model is good as full model

H1: Reduced model is not good as full mode.

Model 1: Y2 ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9

Model 2: Y2 ~ X1 + X2 + X3 + X5 + X6 + X8 + X9

	Df	LogLik	Df	Chisq	Pr(>Chisq)
1	39	-203.26			
2	31	-209.89	-8	13.243	0.1038

Thus , we do not reject H0 and conclude that Reduced model is good as full model.

- **Goodness-of-Fit Tests:**

Hypothesis:H_0: The model fits the data well.

H_1: The model does not fit data well

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	513.110	665	1.000
Deviance	359.465	665	1.000

Interpretation-P-value = 1.0 > 0.05.Therefore, we do not reject Null hypothesis and say that the model fits the data well.

For full model :

➢Training accuracy

Y2 * Predicted Response Category Crosstabulation							
Count							
		Predicted Response Category					Total
		1	2	3	4	5	
Y2	1	3	1	2	0	0	6
	2	0	6	6	1	0	13
	3	0	2	55	11	6	74
	4	0	0	20	35	13	68
	5	0	0	1	16	56	73
Total		3	9	84	63	75	234

Training Accuracy = 66.24%

➢Testing accuracy

Y2 * Predicted Response Category Crosstabulation							
Count							
		Predicted Response Category					Total
		1	2	3	4	5	
Y2	1	0	0	0	0	0	0
	2	0	1	0	0	0	1
	3	0	0	5	7	0	12
	4	0	0	7	8	6	21
	5	0	0	2	4	18	24
Total		0	1	14	19	24	58

Testing Accuracy = 55.17%

For reduced model-

➢Training Accuracy

Y2 * Predicted Response Category Crosstabulation							
Count							
		Predicted Response Category					Total
		1	2	3	4	5	
Y2	1	3	0	0	0	0	3
	2	1	6	2	0	0	9
	3	2	6	55	20	1	84
	4	0	1	11	35	16	63
	5	0	0	6	13	56	75
Total		6	13	74	68	73	234

Training accuracy=64.50%

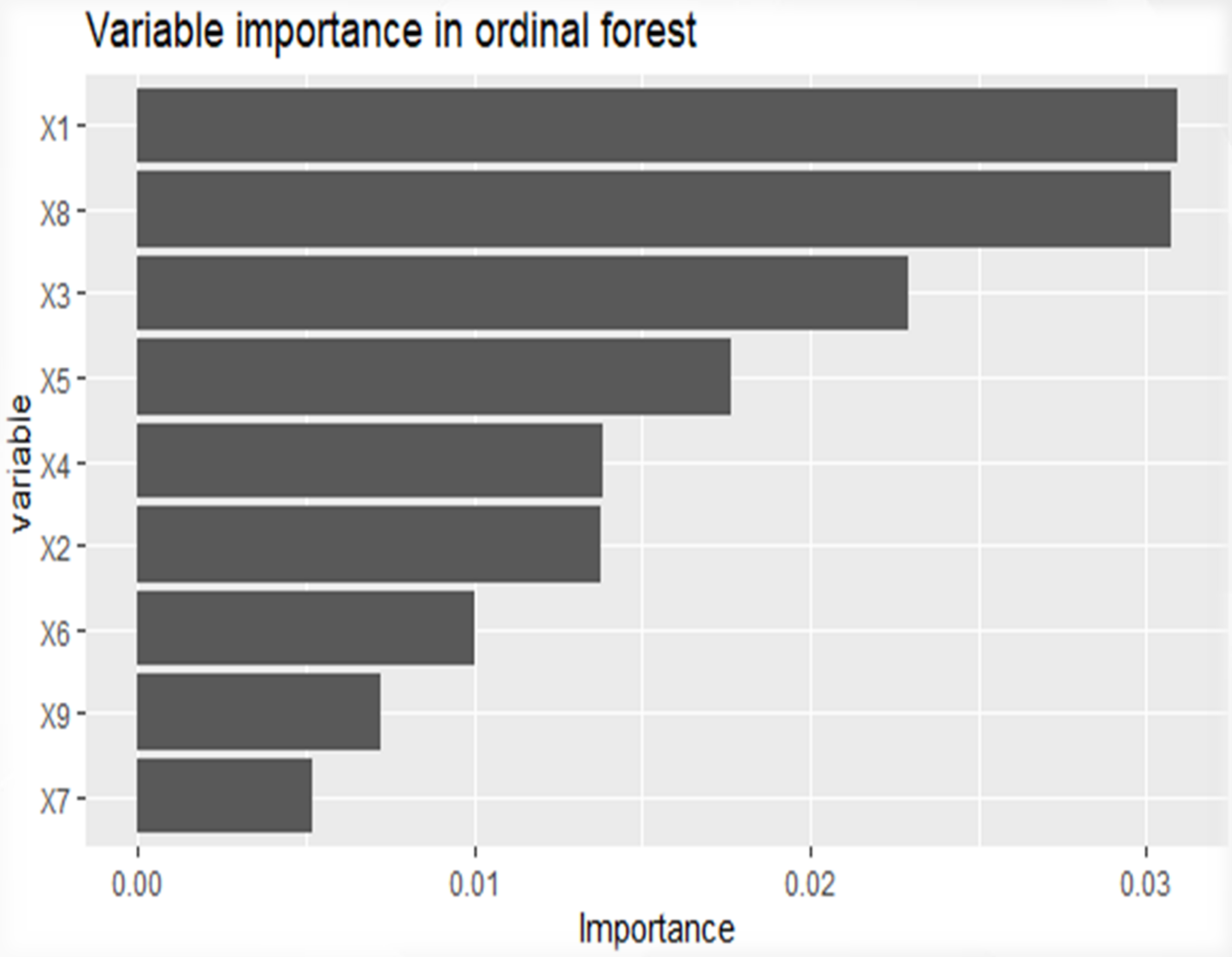
➢Testing Accuracy

Y2 * Predicted Response Category Crosstabulation							
Count							
		Predicted Response Category					Total
		1	2	3	4	5	
Y2	1	0	0	0	0	0	0
	2	0	1	0	0	0	1
	3	0	0	5	7	0	12
	4	0	0	7	8	6	21
	5	0	0	2	4	18	24
Total		0	1	14	19	24	58

Testing accuracy=51.75%

➢ Analysis by Ordinal Forest technique

variable	Importance
X1	X1 0.030935272
X2	X2 0.013744850
X3	X3 0.022907326
X4	X4 0.013857357
X5	X5 0.017680400
X6	X6 0.010049852
X7	X7 0.005210417
X8	X8 0.030759131
X9	X9 0.007252878



Interpretation-From the table of importance and graph we can see that the variable X1,X8,X3 has the most importance compared to rest of the variables.

To test Accuracy

Training the accuracy of the fitted model

pred_labels	1	2	3	4	5
1	3	0	0	0	0
2	1	10	1	0	0
3	0	2	46	1	0
4	0	1	22	58	8
5	0	0	0	4	77

Accuracy=0.8290598

Testing the accuracy of the fitted model

pred_labels	1	2	3	4	5
1	1	0	0	0	0
2	0	0	0	0	0
3	0	0	12	4	0
4	1	1	7	16	6
5	0	0	0	4	6

Accuracy=0.6034483

Training accuracy =82.90%

Testing accuracy =60.34%

Method	Model	Accuracy
Ordinal Logistic Regression	Full Model(Trained Data)	66.24%
	Full Model(Testing Data)	55.17%
	Reduced Model(Trained Data)	64.50%
	Full Model(Testing Data)	51.75%
Ordinal Forest	Full Model (Trained Data)	82.90%
	Full Model(Testing data)	60.34%

<u>Conclusion For Objective 2</u>-

From the best fitted model, we get 7 variables significant,i.e. X1, X2, X3, X5, X6, X8, X9 which are Resource management, strategic sourcing, labour cost management, customer relationship management, Supply chain management, Energy efficiency, product design.

From this data analysis, we get know that the impact of AI on Expenditure is level 4 ,i.e. HIGH

From the ordinal forest technique the accuracy is 82.90% on training data and 60.34% on testing data which gives better accuracy, hence we predict our result from the model fitted by ordinal forest technique which gives us the resulting level 4,i.e HIGH by taking the mode of the predictions. Hence,the impact of AI on Expenditure in industries is HIGH.

Objective 3

To study the impact of Artificial Intelligence on Customer Satisfaction in Industries.

Variables used :-

✓Outcome Variable (Y)– Customer Satisfaction

In objective 2,the **Outcome Variable** as well as the **Predictor Variables** are fitted on **Ordinal Scale** having following levels –

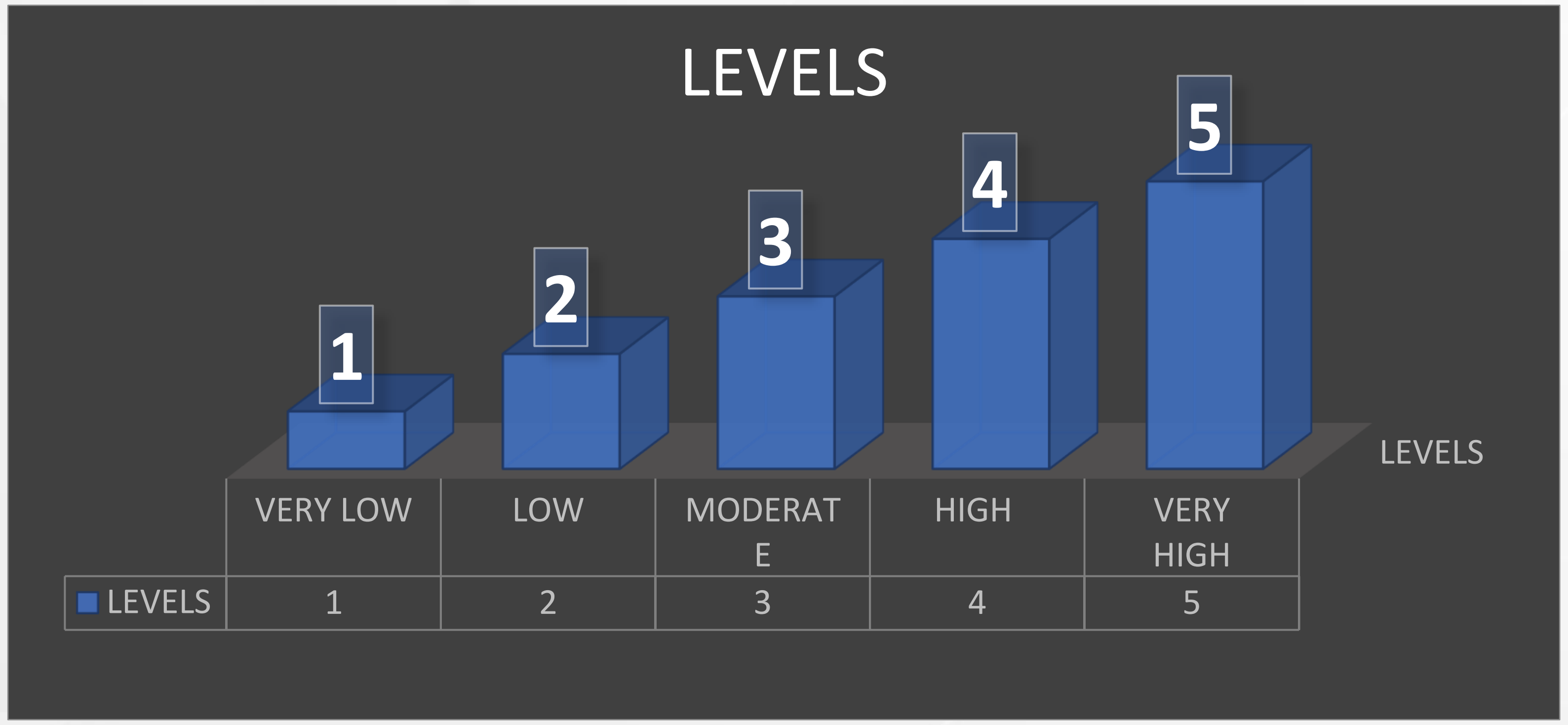
1-Very low

2-Low

3-Moderate

4-High

5-Very high



✓Predictor Variables

1) X1- Product quality
2) X2-Communication with Customer
3) X3-Accessibility
4) X4-Timeliness
5) X5-Feedback
6) X6-Cultural and demographic factors
7) X7-Emotional connection

Analysis for 3rd Objective

- **(Full model Fitting)**
- **Parameter Estimates –**

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Y3 = 1]	-11.827	1.149	106.045	1	<.001	-14.078	-9.576
	[Y3 = 2]	-9.468	.809	136.871	1	<.001	-11.054	-7.882
	[Y3 = 3]	-4.792	.597	64.520	1	<.001	-5.961	-3.623
	[Y3 = 4]	-2.449	.524	21.871	1	<.001	-3.476	-1.423
Location	[X1=1]	-2.430	1.861	1.705	1	.192	-6.077	1.217
	[X1=2]	-2.874	.859	11.193	1	<.001	-4.558	-1.190
	[X1=3]	-1.298	.425	9.301	1	.002	-2.132	-.464
	[X1=4]	-.834	.378	4.862	1	.027	-1.576	-.093
	[X1=5]	0[a]	.	.	0	.	.	.
	[X2=1]	-.092	1.537	.004	1	.952	-3.103	2.920
	[X2=2]	-1.323	.711	3.462	1	.063	-2.717	.071
	[X2=3]	-.332	.446	.555	1	.456	-1.206	.542
	[X2=4]	.239	.376	.405	1	.524	-.498	.976
	[X2=5]	0[a]	.	.	0	.	.	.
	[X3=1]	-1.690	1.229	1.891	1	.169	-4.099	.719
	[X3=2]	-.163	.811	.040	1	.841	-1.752	1.427
	[X3=3]	-.231	.442	.271	1	.602	-1.098	.637
	[X3=4]	.167	.374	.199	1	.655	-.565	.899
	[X3=5]	0[a]	.	.	0	.	.	.
	[X4=1]	3.246	1.463	4.924	1	.026	.379	6.113
	[X4=2]	.158	.835	.036	1	.849	-1.477	1.794
	[X4=3]	-.706	.404	3.061	1	.080	-1.498	.085
	[X4=4]	-.266	.361	.541	1	.462	-.973	.442
	[X4=5]	0[a]	.	.	0	.	.	.
	[X5=1]	-3.399	1.479	5.286	1	.021	-6.297	-.501
	[X5=2]	-2.389	.734	10.582	1	.001	-3.828	-.950
	[X5=3]	-.931	.475	3.836	1	.050	-1.862	.001
	[X5=4]	-.531	.384	1.914	1	.167	-1.283	.221
	[X5=5]	0[a]	.	.	0	.	.	.
	[X6=1]	-2.744	.971	7.991	1	.005	-4.647	-.842
	[X6=2]	-2.799	.691	16.404	1	<.001	-4.153	-1.444
	[X6=3]	-1.684	.466	13.051	1	<.001	-2.598	-.770
	[X6=4]	-1.068	.416	6.601	1	.010	-1.883	-.253
	[X6=5]	0[a]	.	.	0	.	.	.
	[X7=1]	-.992	.670	2.195	1	.138	-2.304	.320
	[X7=2]	-.087	.503	.030	1	.863	-1.072	.899
	[X7=3]	-.734	.424	2.994	1	.084	-1.565	.097
	[X7=4]	-.506	.379	1.783	1	.182	-1.248	.237
	[X7=5]	0[a]	.	.	0	.	.	.

Link function: Logit
(Out of the five levels one level is taken as the reference category ,here the level 5 ,i.e. very high is taken as the reference category)

From full model we get 4 significant variables
i.e. X1, X4, X5, X6.
X1- Product quality
X4-Timeliness
X5-Feedback
X6-Cultural and demographic factors

➢Proportion Odds assumption -

Test of Parallel Lines				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	451.258			
General	406.726[b]	44.532[c]	84	1.000

Hypothesis-H_0: $\beta_{1l} = \beta_1 \quad \forall \quad l = 1, 2, 3, 4$

H_1: $\beta_{1l} \neq \beta_1$ for atleast one l

We Hypothesis that there is a common slope parameter for each of the cumulative logit regression equations instead of 4 distinct slopes.

Interpretation-P-value = 1.00 > 0.05

Therefore, we do not reject the Null hypothesis and conclude that each outcome variable has a common slope parameter.

➢To Check Multicollinearity –

	GVIF	Df	GVIF^(1/(2*Df))
Y1	2.275176	4	1.108222
X1	2.571435	1	1.603569
X2	2.023901	1	1.422639
X3	2.221589	1	1.490500
X4	1.803771	1	1.343046
X5	2.037464	1	1.427398
X6	1.674401	1	1.293987
X7	1.848889	1	1.359738
X8	1.932334	1	1.390084
X9	1.921527	1	1.386191

GVIF^(1/(2*df)) close to 1: Indicates little to no multicollinearity.

GVIF^(1/(2*df)) greater than 1: Indicates the presence of multicollinearity, with higher values suggesting more severe multicollinearity.

Interpretation: In the above table all the adjusted GVIF values are close to 1 which indicates relatively little to no collinearity which is acceptable.

❖ **(Reduced model fitting)**

➢ **Parameter estimates**

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound
Threshold	[Y3 = 1]	-11.155	1.059	111.009	1	<.001	-13.231
	[Y3 = 2]	-8.911	.737	146.042	1	<.001	-10.356
	[Y3 = 3]	-4.477	.524	72.887	1	<.001	-5.505
	[Y3 = 4]	-2.278	.457	24.856	1	<.001	-3.173
Location	[X1=1]	-1.814	1.685	1.159	1	.282	-5.117
	[X1=2]	-2.666	.734	13.188	1	<.001	-4.104
	[X1=3]	-1.363	.407	11.219	1	<.001	-2.160
	[X1=4]	-.882	.361	5.962	1	.015	-1.589
	[X1=5]	0[a]	.	.	0	.	.
	[X4=1]	2.600	1.289	4.068	1	.044	.074
	[X4=2]	-.429	.732	.343	1	.558	-1.863
	[X4=3]	-.914	.383	5.681	1	.017	-1.665
	[X4=4]	-.252	.345	.536	1	.464	-.927
	[X4=5]	0[a]	.	.	0	.	.
	[X5=1]	-4.100	1.395	8.637	1	.003	-6.834
	[X5=2]	-2.736	.653	17.565	1	<.001	-4.016
	[X5=3]	-1.283	.432	8.803	1	.003	-2.130
	[X5=4]	-.569	.363	2.466	1	.116	-1.280
	[X5=5]	0[a]	.	.	0	.	.
	[X6=1]	-3.416	.881	15.025	1	<.001	-5.144
	[X6=2]	-3.100	.577	28.817	1	<.001	-4.232
	[X6=3]	-1.766	.404	19.076	1	<.001	-2.558
	[X6=4]	-.971	.389	6.222	1	.013	-1.733
	[X6=5]	0[a]	.	.	0	.	.

Fitted Ordinal logistic model-

$$\Pi_1 = \frac{e^{logit\ F1}}{1+e^{logitF1}}$$

Where,

logitF1=-11.155-1.814*Product Quality(1)-2.666*Product Quality(2)-1.363*Product Quality(3)-0.882*Product Quality(4)+2.600*Timeliness(1)-0.429*Timeliness(2)-0.914*Timeliness(3)-0.252*Timeliness(4)-4.1*Feedback(1)-2.736*Feedback(2)-1.283*Feedback(3)-0.569*Feedback(4)-3.416*Cultural and Demographic Factors(1)-3.100*Cultural and Demographic Factors(2)-1.766*Cultural and Demographic Factors(3)-0.971*Cultural and Demographic Factors(4)

$$\Pi_2 = \frac{e^{logit\ F2}}{1+e^{logitF2}}$$

Where,

logitF2=-8.911-1.814*Product Quality(1)-2.666*Product Quality(2)-1.363*Product Quality(3)-0.882*Product Quality(4)+2.600*Timeliness(1)-0.429*Timeliness(2)-0.914*Timeliness(3)-0.252*Timeliness(4)-4.1*Feedback(1)-2.736*Feedback(2)-1.283*Feedback(3)-0.569*Feedback(4)-3.416*Cultural and Demographic Factors(1)-3.100*Cultural and Demographic Factors(2)-1.766*Cultural and Demographic Factors(3)-0.971*Cultural and Demographic Factors(4)

$$\Pi_3 = \frac{e^{logit\ F3}}{1+e^{logitF3}}$$

Where,

logitF3=-4.477-1.814*Product Quality(1)-2.666*Product Quality(2)-1.363*Product Quality(3)-0.882*Product Quality(4)+2.600*Timeliness(1)-0.429*Timeliness(2)-0.914*Timeliness(3)-0.252*Timeliness(4)-4.1*Feedback(1)-2.736*Feedback(2)-1.283*Feedback(3)-0.569*Feedback(4)-3.416*Cultural and Demographic Factors(1)-3.100*Cultural and Demographic Factors(2)-1.766*Cultural and Demographic Factors(3)-0.971*Cultural and Demographic Factors(4)

$$\Pi_4 = \frac{e^{logit\ F4}}{1+e^{logitF4}}$$

Where,

logitF4=-2.278-1.814*Product Quality(1)-2.666*Product Quality(2)-1.363*Product Quality(3)-0.882*Product Quality(4)+2.600*Timeliness(1)-0.429*Timeliness(2)-0.914*Timeliness(3)-0.252*Timeliness(4)-4.1*Feedback(1)-2.736*Feedback(2)-1.283*Feedback(3)-0.569*Feedback(4)-3.416*Cultural and Demographic Factors(1)-3.100*Cultural and Demographic Factors(2)-1.766*Cultural and Demographic Factors(3)-0.971*Cultural and Demographic Factors(4)

Proportion Of odds Assumption-

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	342.970			
General	330.133[b]	12.837[c]	48	1.000

We test,

H0: $\beta_{1l} = \beta_1$ $i.e.$ $\beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = \beta_1 \quad \forall \quad l = 1, 2, \dots, k-1$

H1: $\beta_{1l} \neq \beta_1$ for atleast one l

Where β_1 depicts common slope parameter.

We Hypothesis that there is common slope parameter

Interpretation-P-value = 1.00 > 0.05

Therefore, we do not reject Null hypothesis and conclude that there is common slope parameter for each outcome variable.

Model fitting information

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	565.651			
Final	342.970	222.680	16	<.001

Hypothesis:

H_0: The predictor variables do not significantly improve the fit of the model compared to the null model (without predictors)

H_1: At least one of the predictor variables significantly improves the fit of the model

Interpretation: P-value = 0.001 < 0.05 of Final Model

Therefore, we reject the Null hypothesis and say that the predictor variables significantly improve the fit of the model.

Likelihood Ratio Test –

Model 1: Y3 ~ X1 + X2 + X3 + X4 + X5 + X6 + X7

Model 2: Y3 ~ X1 + X4 + X5 + X6

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	44	-212			
2	32	-216.90	-12.	9.1206	0.6926

Interpretation: P-value =0.6926 > 0.05

Hence, we do not reject the null hypothesis and conclude that the Reduced model is as good as the Full model.

Goodness-of-Fit Tests:

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	517.408	360	1.000
Deviance	262.181	360	1.000

Hypothesis: H_0: The model fits the data well.

H_1: The model does not fit the data well

Interpretation – P-value – 1.000 >0.05. Therefore, we do not reject the Null hypothesis and say that the model fits the data well.

➢For full model

Confusion matrix –

Y3 * Predicted Response Category Crosstabulation

Count		Predicted Response Category					Total
		1	2	3	4	5	
Y3	1	0	0	2	0	0	2
	2	1	1	8	0	0	10
	3	0	0	73	16	8	97
	4	0	0	17	48	22	87
	5	0	0	2	20	74	96
Total		1	1	102	84	104	292

Accuracy of the full model is 67.12%.

➢ For reduced model

Y3 * Predicted Response Category Crosstabulation

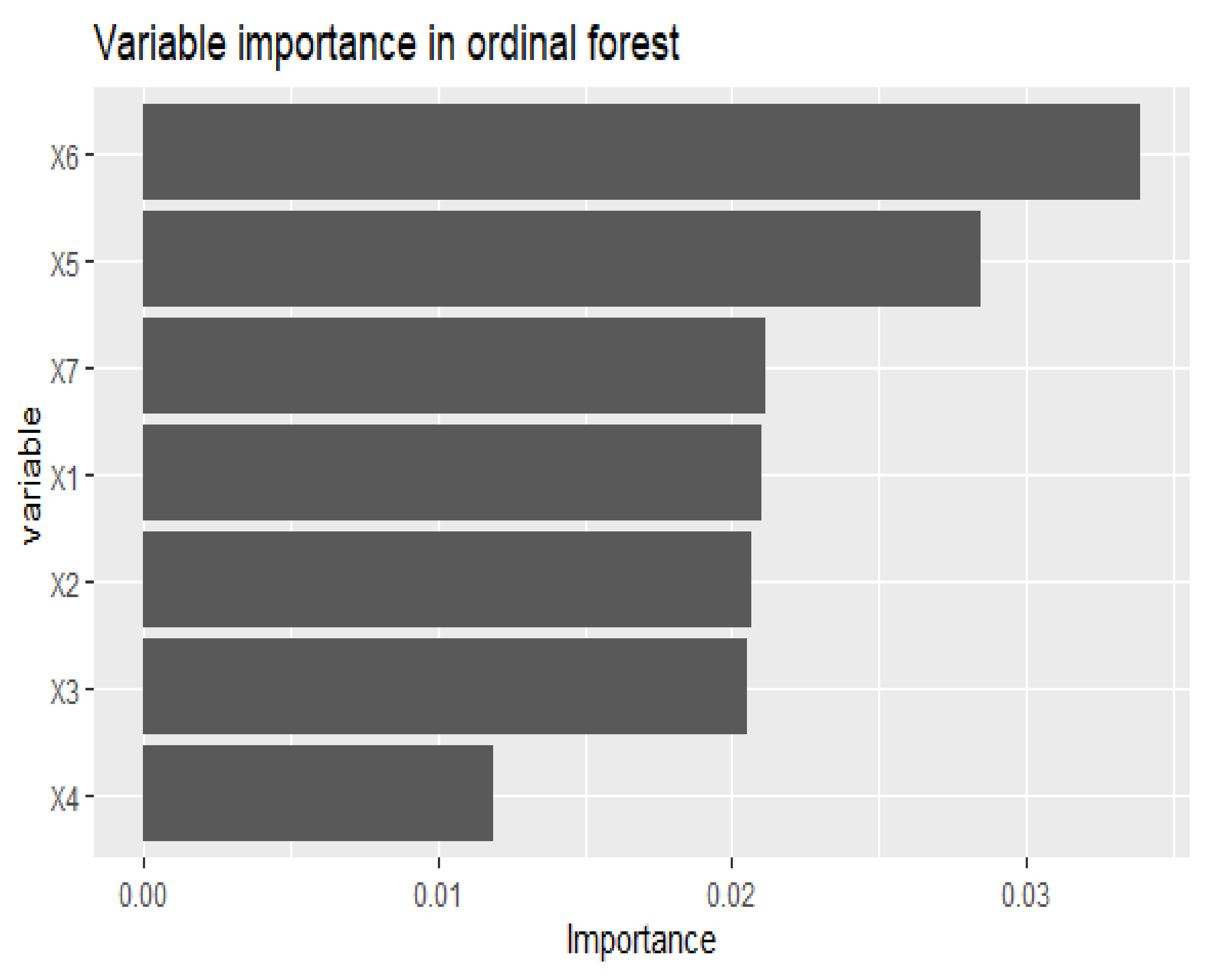
Count		Predicted Response Category					Total
		1	2	3	4	5	
Y3	1	0	0	2	0	0	2
	2	1	0	9	0	0	10
	3	0	1	68	20	8	97
	4	0	0	17	49	21	87
	5	0	0	1	23	72	96
Total		1	1	97	92	101	292

Accuracy of the reduced model is 64.73%.

➢ Analysis by Ordinal Forest technique

variable	Importance
x1	x1 0.02100203
x2	x2 0.02069795
x3	x3 0.02052550
x4	x4 0.01192263
x5	x5 0.02847894
x6	x6 0.03390493
x7	x7 0.02113547

Importance of variables given by ordinal forest technique



<u>Interpretation</u>-From the table of importance and graph, we can see that the variables X6 & X5 have the most importance as compared to the rest of the variables.

❖ **Training accuracy of the above fitted model**

Confusion matrix

pred_labels	1	2	3	4	5
1	0	0	0	0	0
2	2	3	2	0	0
3	0	6	78	15	0
4	0	0	0	36	0
5	0	0	8	24	82
Accuracy=0.7773438					

Training accuracy = **77.73%** .

❖ **Testing accuracy of the above fitted model**

Confusion matrix

pred_labels	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	0	0
3	0	1	7	4	1
4	0	0	0	2	0
5	0	0	2	6	13
Accuracy=0.6111111					

Testing accuracy = **61.11**%.

➢ Accuracy table

Method	Model	Accuracy
Ordinal Logistic Regression	Full Model	67.12%
	Reduced Model	64.73%
Ordinal Forest	Full Model (Trained Data)	77.73%
	Full Model(Testing data)	61.11%

Conclusion for Objective 3-
So from the best-fitted model,i.e.reduced model we get 7 significant variables,i.e. X1,X4,X5,X6 which are product quality, timeliness, feedback, cultural and demographic factors
From this data analysis, we come to know that the impact of AI on Customer Satisfaction is level 5 ,i.e. VERY HIGH
Similarly, objectives 1 & 2 , to increase the accuracy of the model we used machine learning technique,i.e ordinal forest technique which gives us an accuracy of 77.73% on training data and 61.11% on testing data which gives better accuracy,hence we predict our result from the model fitted by ordinal forest technique which gives us the resulting level 5,i.e VERY HIGH by taking the mode of the predictions. Hence,**the impact of AI on Customer Satisfaction in industries is VERY HIGH.**
Due to not having sufficient data this problem may have been occurred , so , additionally, gathering more data could further improve the model's performance and robustness. For further, use method like cross validation, bagging which can contribute to reducing overfitting by simplifying the model.

Thank You