

Inhaltsverzeichnis

Angaben zur Firma	i
Eidesstattliche Erklärung	ii
Kurzfassung	iii
Inhaltsverzeichnis	iv
Abbildungsverzeichnis	vi
Quelltextverzeichnis	xxii
Abkürzungsverzeichnis	xxiii
A. Appendix Binary Classification Experiments	1
B. Appendix Multiclass Classification Experiments	12
C. Appendix Regression Experiments	23
D. Appendix Binary Classification Subset Experiments	34
E. Appendix Multiclass Classification Subset Experiments	45
F. Appendix Regression Subset Experiments	56
G. Appendix Binary Classification Baseline - Imputed Comparisons	67

H. Appendix Binary Classification Imputed - Subset Comparisons	76
I. Appendix Multiclass Classification Baseline - Imputed Comparisons	85
J. Appendix Multiclass Classification Imputed - Subset Comparisons	94
K. Appendix Regression Baseline - Imputed Comparisons	103
L. Appendix Regression Imputed - Subset Comparisons	112
M. Important Changes in the Framework	121

Abbildungsverzeichnis

A.1. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points	1
A.2. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Imputation Methods	2
A.3. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction	2
A.4. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missingness Pattern	3
A.5. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points .	3
A.6. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method	4
A.7. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction	4
A.8. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points Missingness Pattern	5
A.9. Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction	5
A.10. Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction	6
A.11. Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction	6
A.12. Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction	7

A.13.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction	7
A.14.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction	8
A.15.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction	8
A.16.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction	9
A.17.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction	9
A.18.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction	10
A.19.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction	10
A.20.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction	11
B.1. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points	12
B.2. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method	13
B.3. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction	13
B.4. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missingness Pattern	14
B.5. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points .	14
B.6. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method	15
B.7. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction	15

B.8. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Missingness Pattern	16
B.9. Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction	16
B.10.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction	17
B.11.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction	17
B.12.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction	18
B.13.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction	18
B.14.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction	19
B.15.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction	19
B.16.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction	20
B.17.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction	20
B.18.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction	21
B.19.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction	21
B.20.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction	22
C.1. Improvement for Experiments Relative to the Average Best Imputation Method by Percentage	23
C.2. Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Imputation Method	24

C.3. Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Missing Fraction	24
C.4. Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Missingness Pattern	25
C.5. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage	25
C.6. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Imputation Method	26
C.7. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Missing Fraction	26
C.8. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Missingness Pattern	27
C.9. Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction	27
C.10.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction	28
C.11.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction	28
C.12.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction	29
C.13.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction	29
C.14.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction	30
C.15.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction	30
C.16.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction	31
C.17.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction	31

C.18.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction	32
C.19.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction	32
C.20.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction	33
D.1. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points	34
D.2. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Imputation Methods	35
D.3. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction	35
D.4. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missingness Pattern	36
D.5. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points .	36
D.6. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method	37
D.7. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction	37
D.8. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points Missingness Pattern	38
D.9. Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction	38
D.10.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction	39
D.11.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction	39
D.12.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction	40

D.13.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction	40
D.14.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction	41
D.15.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction	41
D.16.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction	42
D.17.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction	42
D.18.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction	43
D.19.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction	43
D.20.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction	44
E.1. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points	45
E.2. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method	46
E.3. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction	46
E.4. Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missingness Pattern	47
E.5. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points .	47
E.6. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method	48
E.7. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction	48

E.8. Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Missingness Pattern	49
E.9. Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction	49
E.10.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction	50
E.11.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction	50
E.12.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction	51
E.13.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction	51
E.14.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction	52
E.15.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction	52
E.16.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction	53
E.17.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction	53
E.18.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction	54
E.19.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction	54
E.20.Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction	55
F.1. Improvement for Experiments Relative to the Average Best Imputation Method by Percentage	56
F.2. Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Imputation Method	57

F.3.	Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Missing Fraction	57
F.4.	Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Missingness Pattern	58
F.5.	Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage	58
F.6.	Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Imputation Method	59
F.7.	Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Missing Fraction	59
F.8.	Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Missingness Pattern	60
F.9.	Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction	60
F.10.	Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction	61
F.11.	Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction	61
F.12.	Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction	62
F.13.	Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction	62
F.14.	Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction	63
F.15.	Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction	63
F.16.	Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction	64
F.17.	Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction	64

F.18. Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction	65
F.19. Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction	65
F.20. Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction	66
G.1. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data	67
G.2. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Imputation Method	68
G.3. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Fraction	68
G.4. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Pattern	69
G.5. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.01 Missing Fraction . . .	69
G.6. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.1 Missing Fraction . . .	70
G.7. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.3 Missing Fraction . . .	70
G.8. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.5 Missing Fraction . . .	71
G.9. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.01 Missing Fraction . .	71
G.10. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.1 Missing Fraction . .	72
G.11. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.3 Missing Fraction . .	72
G.12. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.5 Missing Fraction . .	73
G.13. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.01 Missing Fraction . .	73

G.14.Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.1 Missing Fraction	74
G.15.Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.3 Missing Fraction	74
G.16.Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.5 Missing Fraction	75
H.1. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data	76
H.2. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Imputation Method	77
H.3. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Fraction	77
H.4. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Pattern	78
H.5. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.01 Missing Fraction	78
H.6. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.1 Missing Fraction	79
H.7. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.3 Missing Fraction	79
H.8. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.5 Missing Fraction	80
H.9. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.01 Missing Fraction	80
H.10.Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.1 Missing Fraction	81

H.11	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.3 Missing Fraction	81
H.12	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.5 Missing Fraction	82
H.13	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.01 Missing Fraction	82
H.14	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.1 Missing Fraction	83
H.15	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.3 Missing Fraction	83
H.16	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.5 Missing Fraction	84
I.1.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data	85
I.2.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Imputation Method	86
I.3.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Fraction	86
I.4.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Pattern	87
I.5.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.01 Missing Fraction	87
I.6.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.1 Missing Fraction	88
I.7.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.3 Missing Fraction	88
I.8.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.5 Missing Fraction	89

I.9.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.01 Missing Fraction	89
I.10.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.1 Missing Fraction	90
I.11.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.3 Missing Fraction	90
I.12.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.5 Missing Fraction	91
I.13.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.01 Missing Fraction	91
I.14.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.1 Missing Fraction	92
I.15.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.3 Missing Fraction	92
I.16.	Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.5 Missing Fraction	93
J.1.	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data	94
J.2.	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Imputation Method	95
J.3.	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Fraction	95
J.4.	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Pattern	96
J.5.	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.01 Missing Fraction	96
J.6.	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.1 Missing Fraction	97
J.7.	Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.3 Missing Fraction	97

J.8. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.5 Missing Fraction	98
J.9. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.01 Missing Fraction	98
J.10. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.1 Missing Fraction	99
J.11. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.3 Missing Fraction	99
J.12. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.5 Missing Fraction	100
J.13. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.01 Missing Fraction	100
J.14. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.1 Missing Fraction	101
J.15. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.3 Missing Fraction	101
J.16. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.5 Missing Fraction	102
K.1. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data	103
K.2. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Imputation Method	104
K.3. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Fraction	104

K.4. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Pattern	105
K.5. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.01 Missing Fraction . . .	105
K.6. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.1 Missing Fraction . . .	106
K.7. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.3 Missing Fraction . . .	106
K.8. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.5 Missing Fraction . . .	107
K.9. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.01 Missing Fraction . .	107
K.10. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.1 Missing Fraction . .	108
K.11. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.3 Missing Fraction . .	108
K.12. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.5 Missing Fraction . .	109
K.13. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.01 Missing Fraction . .	109
K.14. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.1 Missing Fraction . .	110
K.15. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.3 Missing Fraction . .	110
K.16. Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.5 Missing Fraction . .	111
L.1. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data	112
L.2. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Imputation Method	113
L.3. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Fraction	113

L.4. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Pattern	114
L.5. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.01 Missing Fraction	114
L.6. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.1 Missing Fraction	115
L.7. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.3 Missing Fraction	115
L.8. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.5 Missing Fraction	116
L.9. Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.01 Missing Fraction	116
L.10.Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.1 Missing Fraction	117
L.11.Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.3 Missing Fraction	117
L.12.Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.5 Missing Fraction	118
L.13.Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.01 Missing Fraction	118
L.14.Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.1 Missing Fraction	119
L.15.Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.3 Missing Fraction	119

L.16.Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.5 Missing Fraction	120
--	-----

Listings

M.1. Preparation for Model Training with Imputed Data	121
M.2. SGDClassifier Random Seed	122
M.3. SGDRegressor Random Seed	122
M.4. Column Selection MAR Experiments	123
M.5. OpenML Random Seed	124
M.6. Subset Experiment Adjustments	125
M.7. Random Seed for Multiple Repetitions of Experiments	126
M.8. Adjustment for the Usage of the Imputed Training and Test Data . . .	127

A. Appendix Binary Classification Experiments

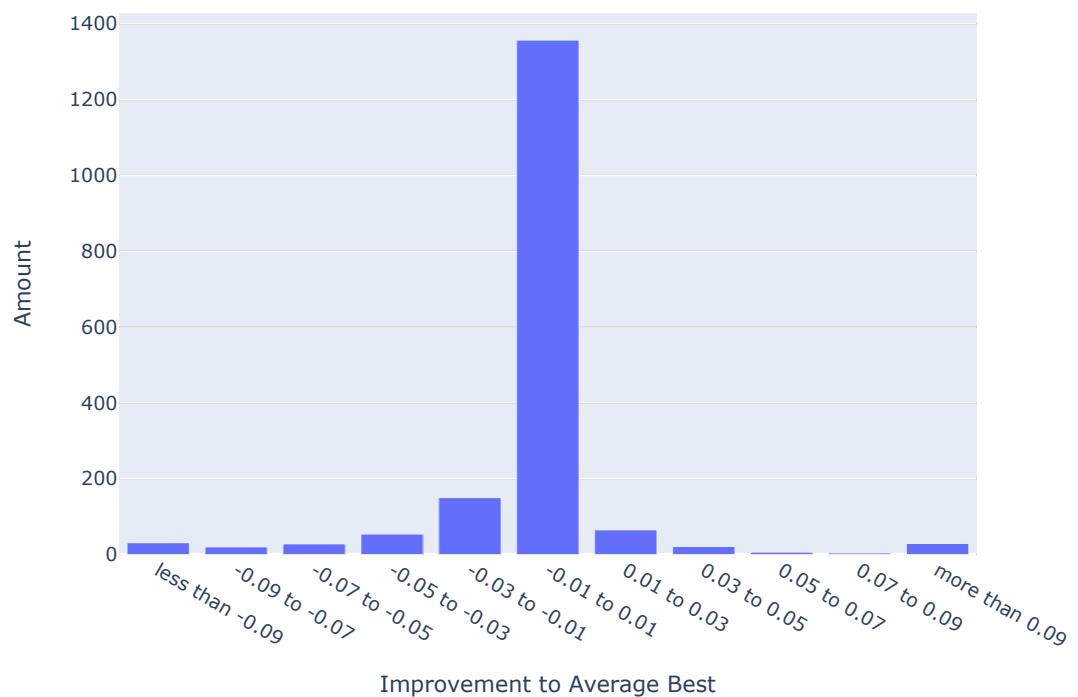


Abbildung A.1.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points

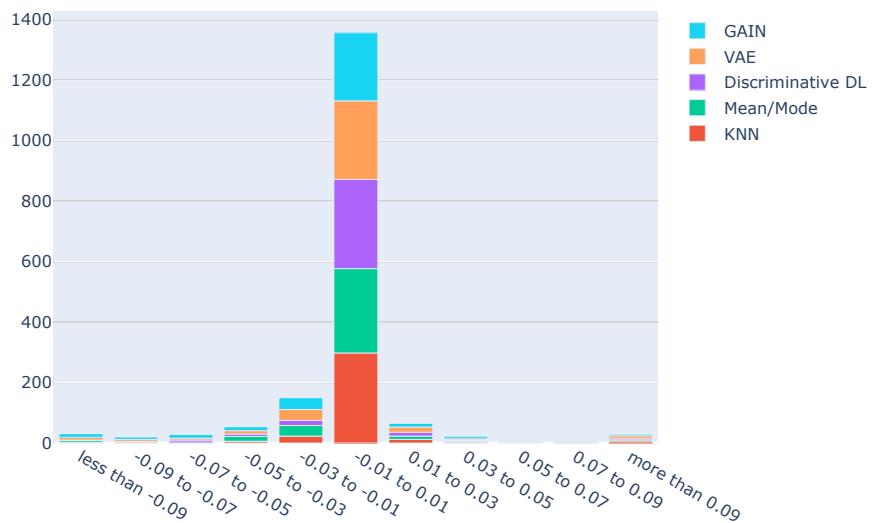


Abbildung A.2.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Imputation Methods

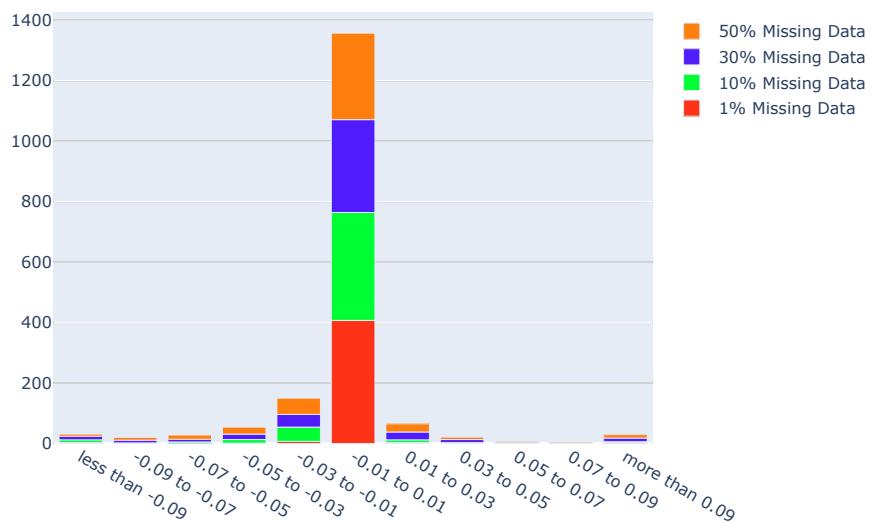


Abbildung A.3.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction

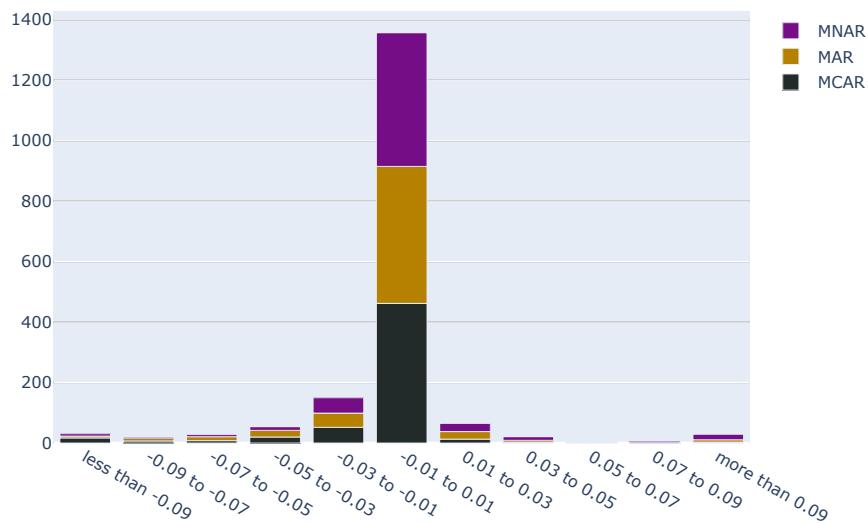


Abbildung A.4.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missingness Pattern

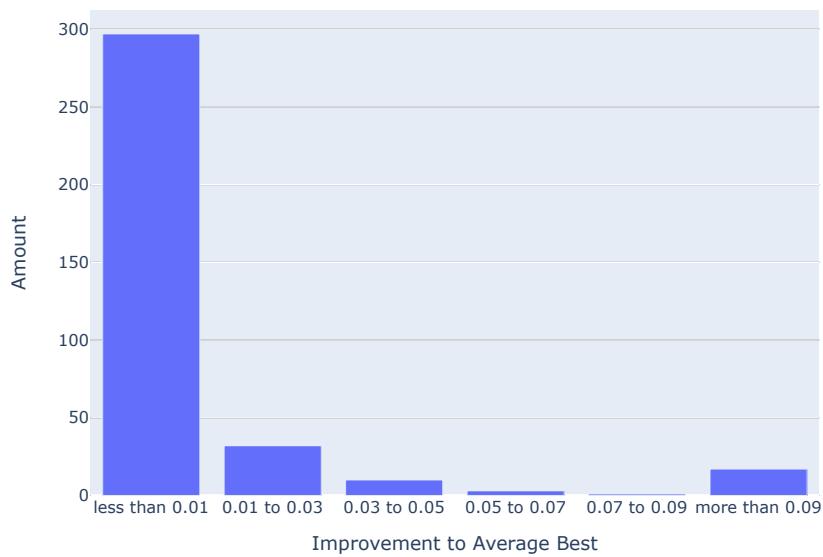


Abbildung A.5.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points

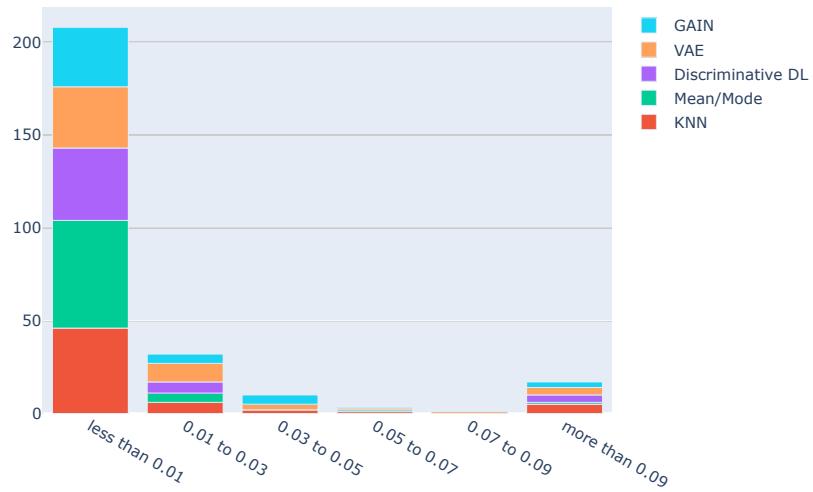


Abbildung A.6.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method

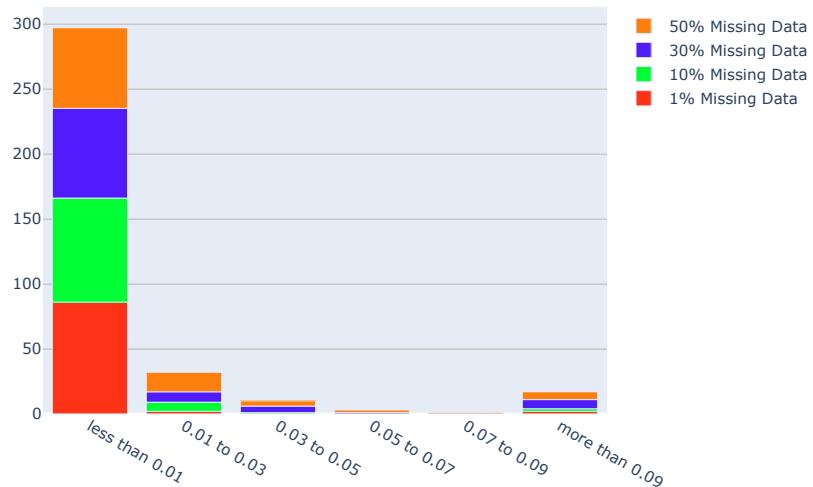


Abbildung A.7.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction

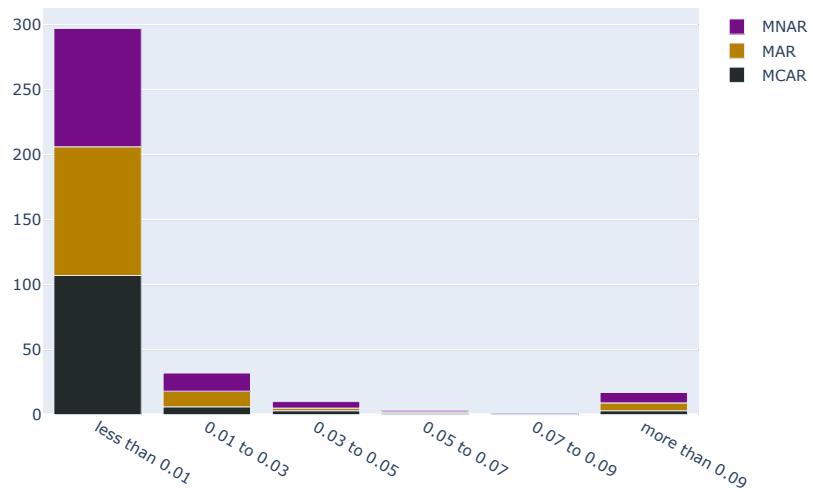


Abbildung A.8.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points Missingness Pattern

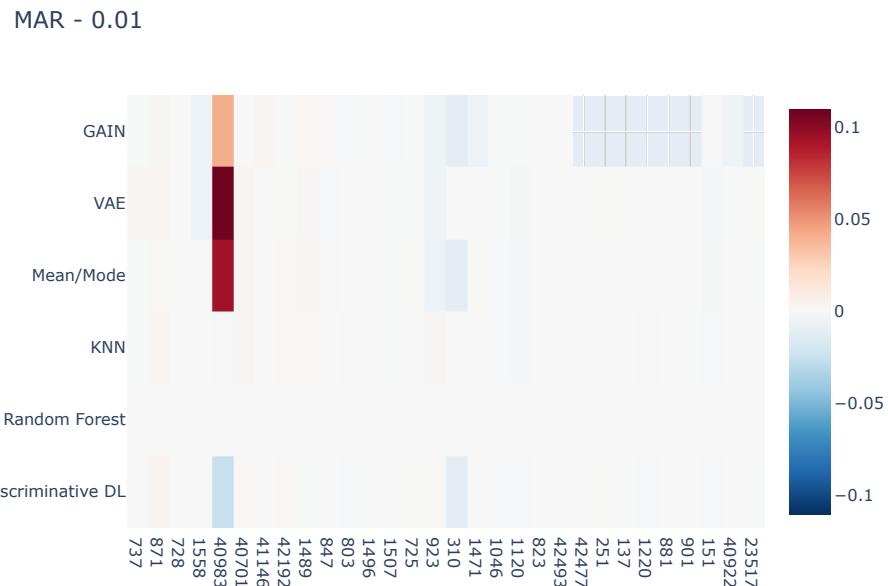


Abbildung A.9.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction

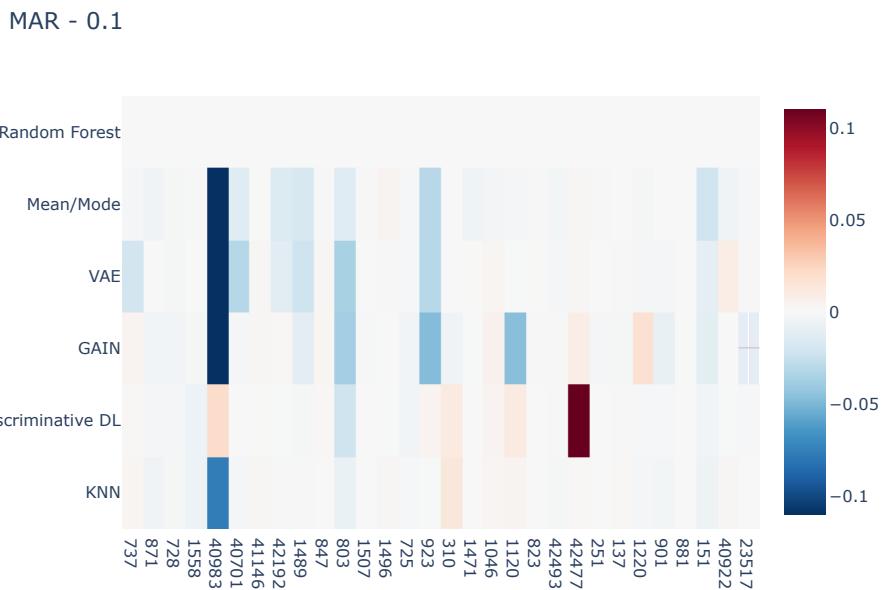


Abbildung A.10.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction

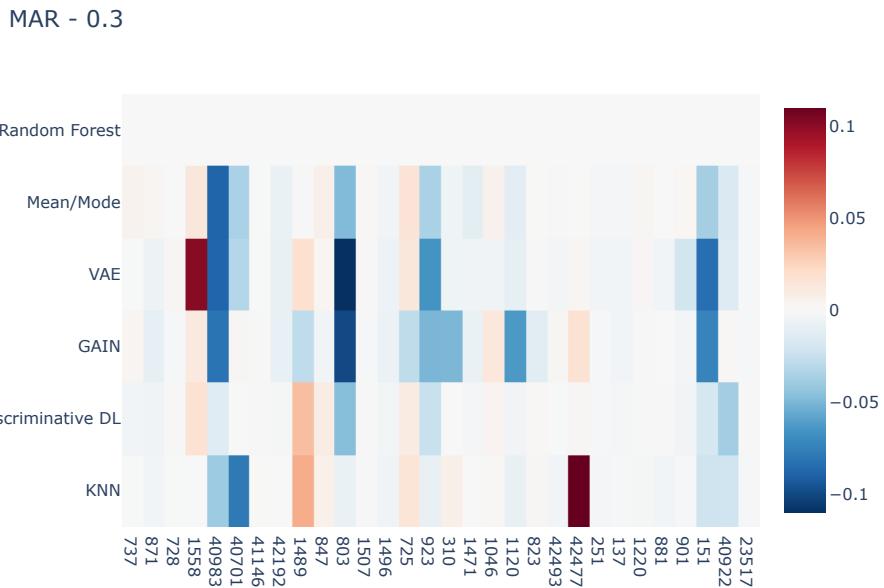


Abbildung A.11.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction

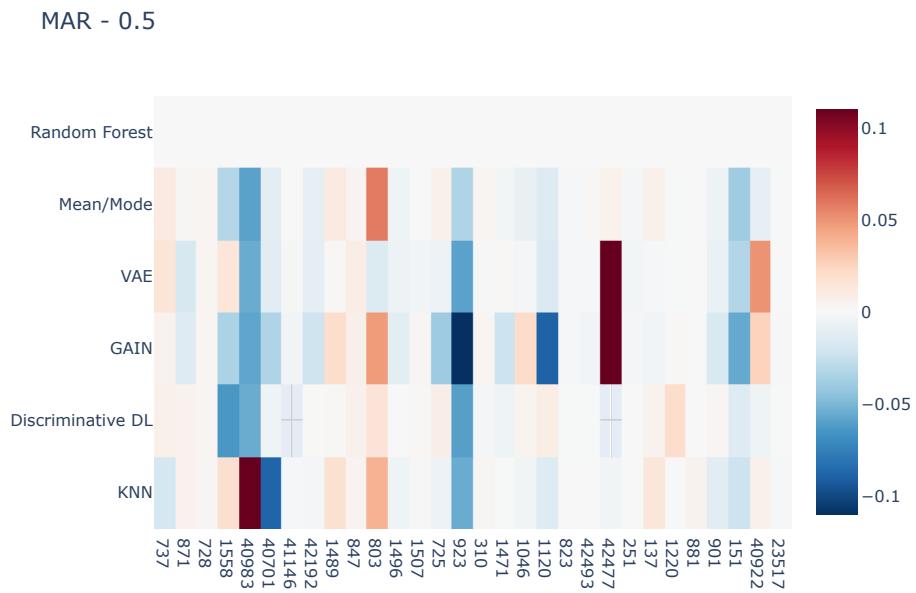


Abbildung A.12.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction

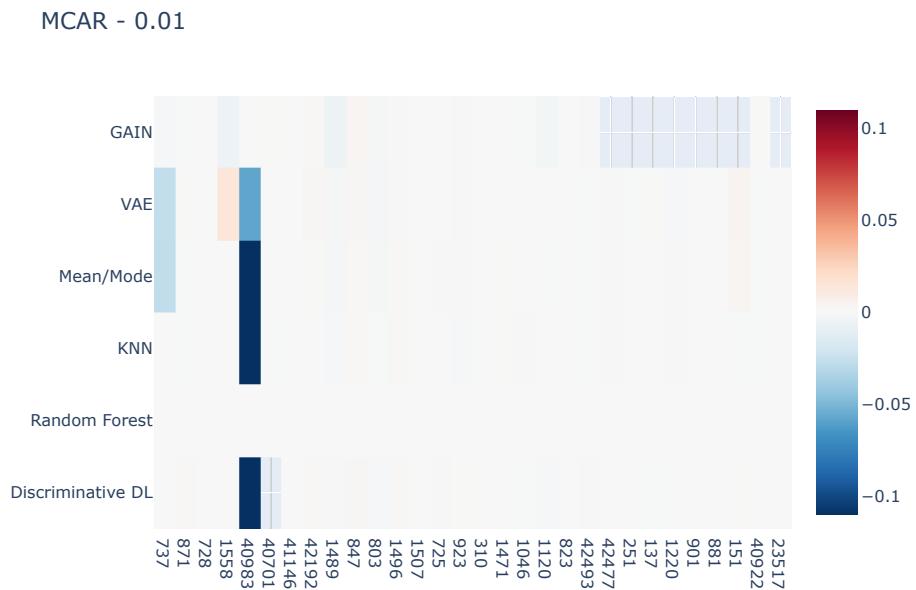


Abbildung A.13.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction

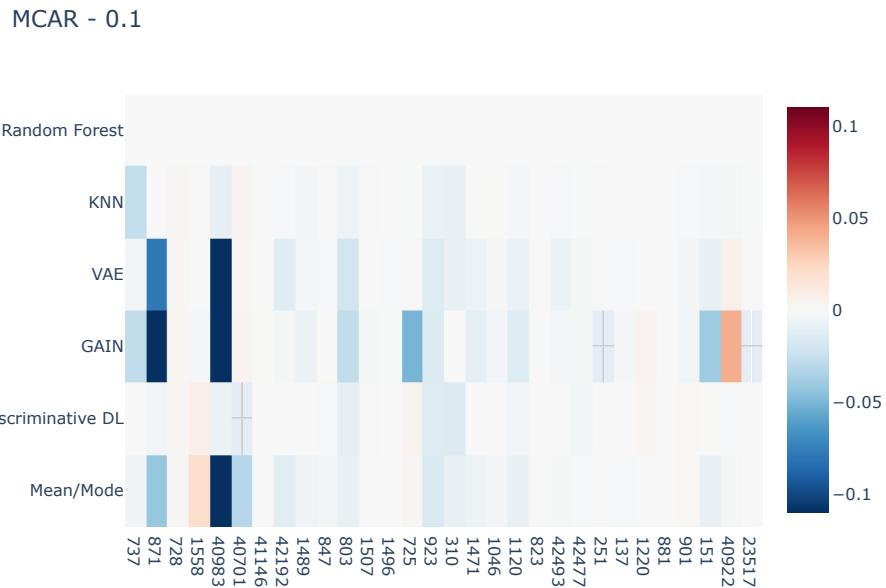


Abbildung A.14.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction

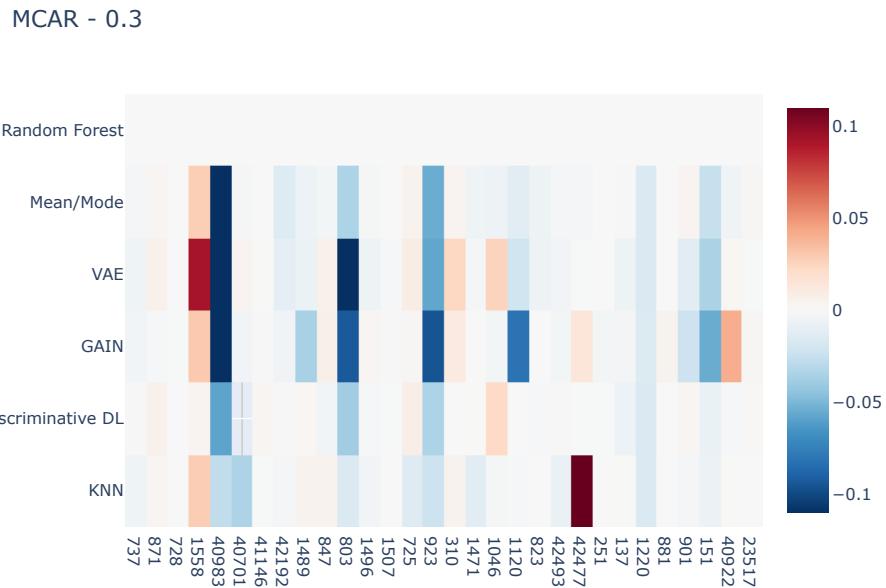


Abbildung A.15.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction

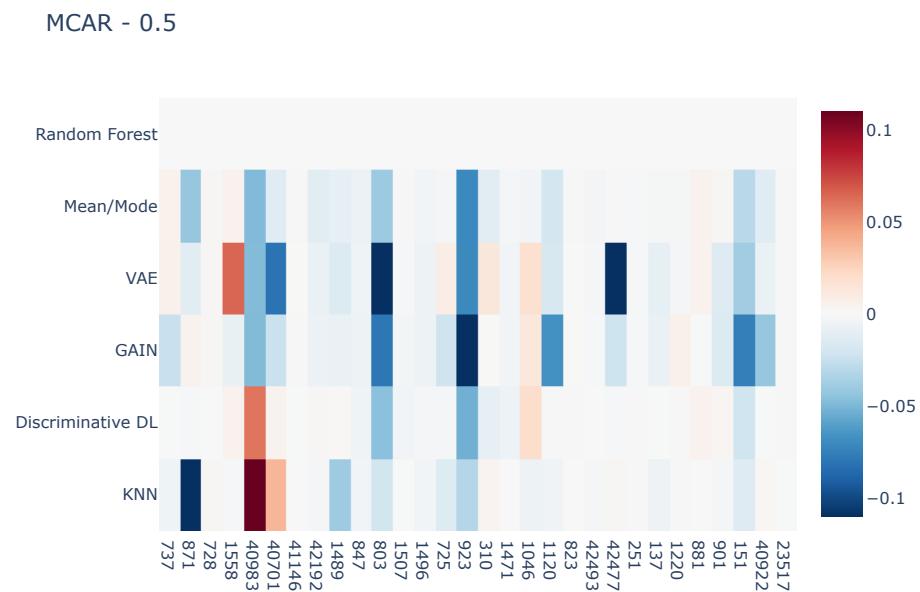


Abbildung A.16.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction

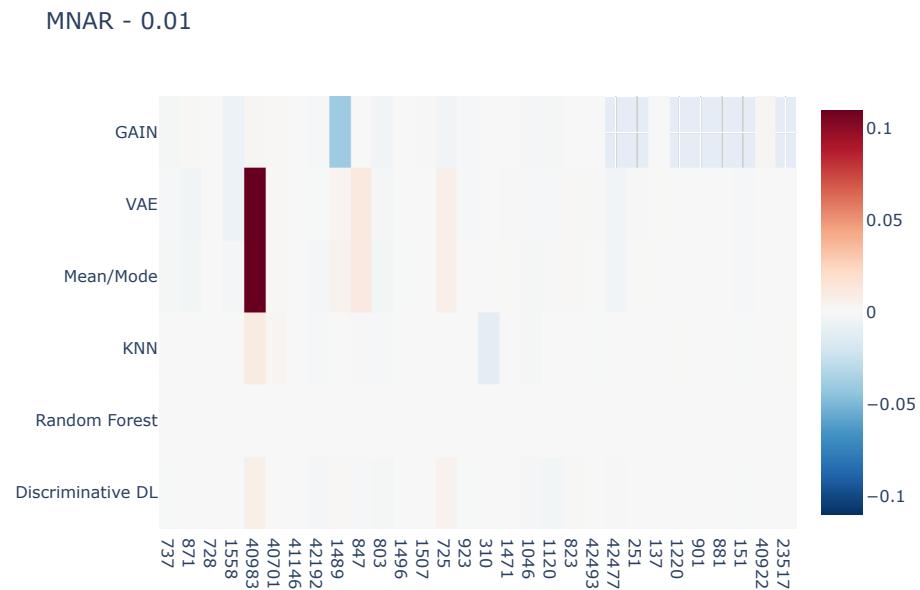


Abbildung A.17.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction

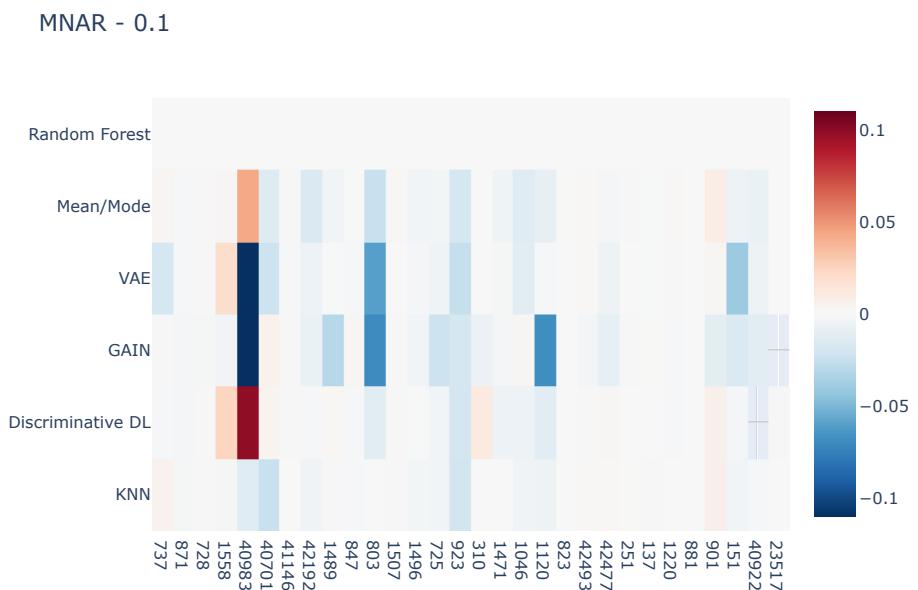


Abbildung A.18.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction

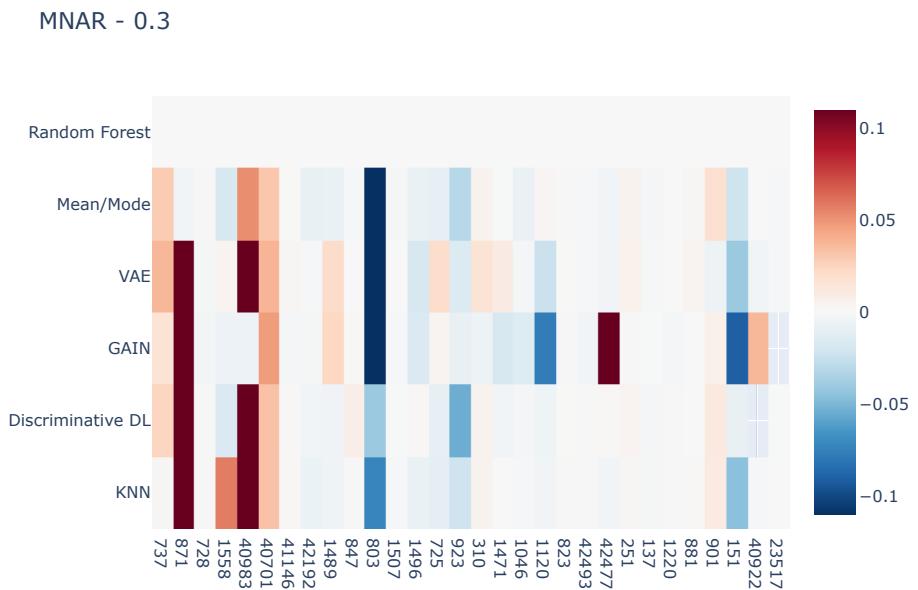


Abbildung A.19.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction

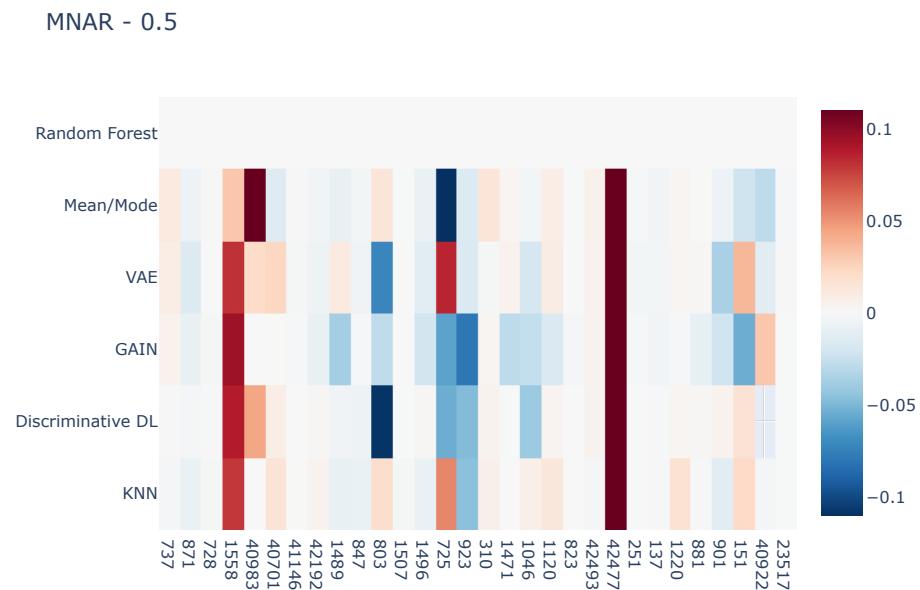


Abbildung A.20.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction

B. Appendix Multiclass Classification Experiments

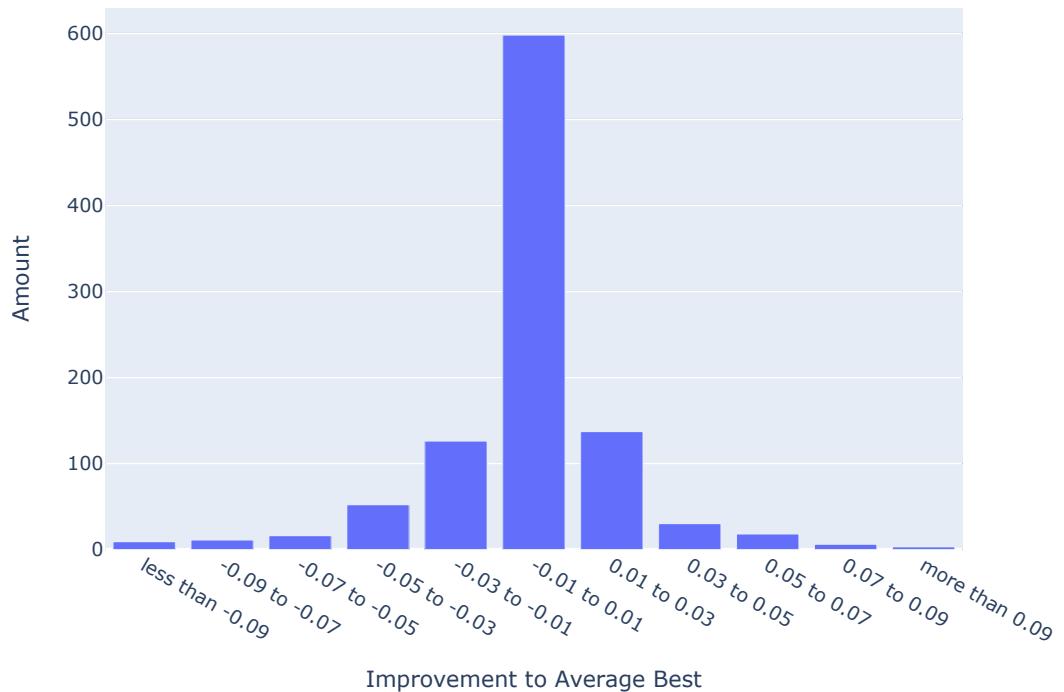


Abbildung B.1.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points

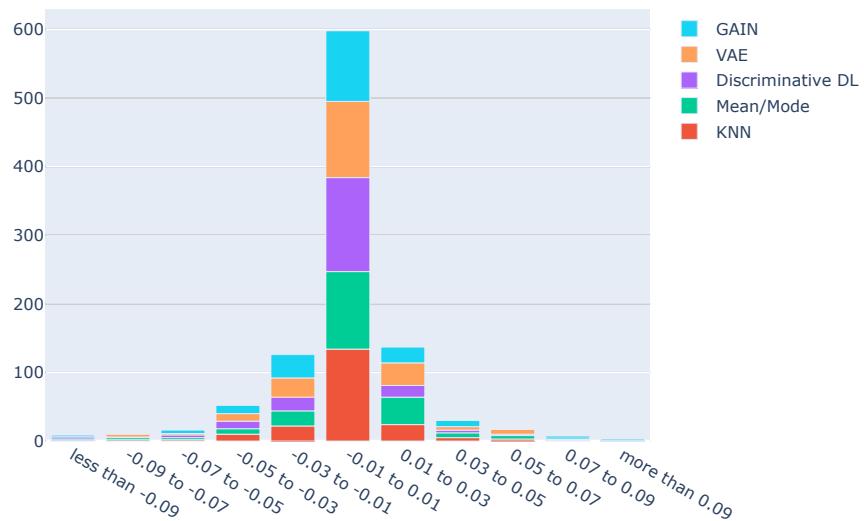


Abbildung B.2.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method

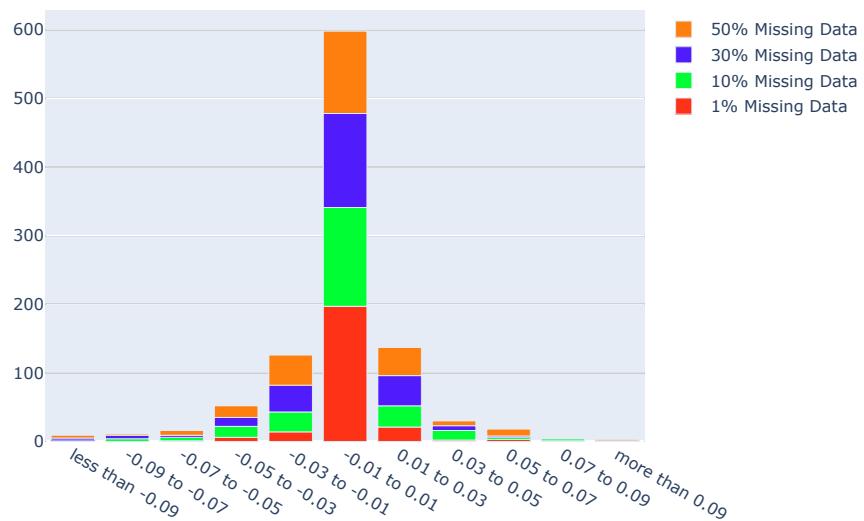


Abbildung B.3.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction

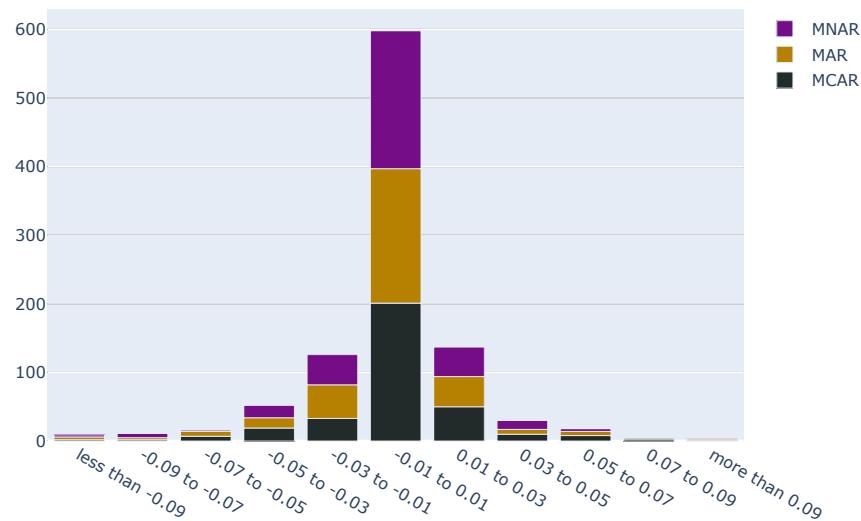


Abbildung B.4.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missingness Pattern

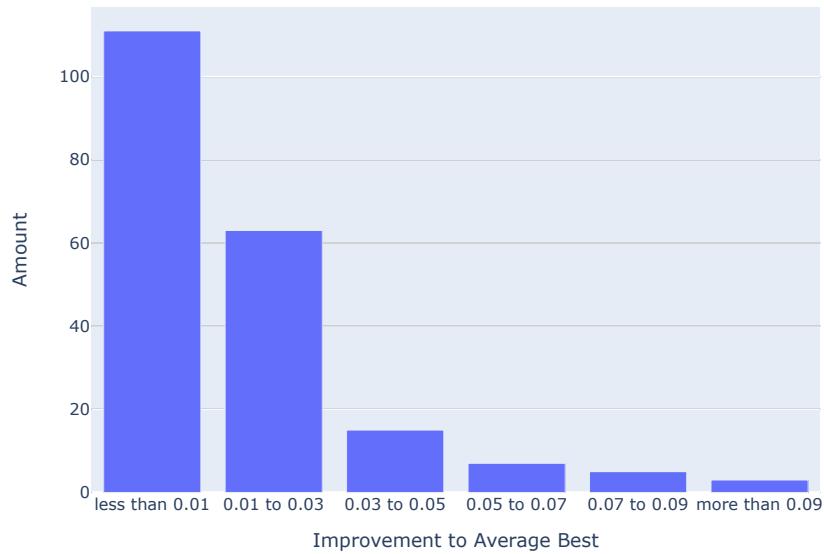


Abbildung B.5.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points

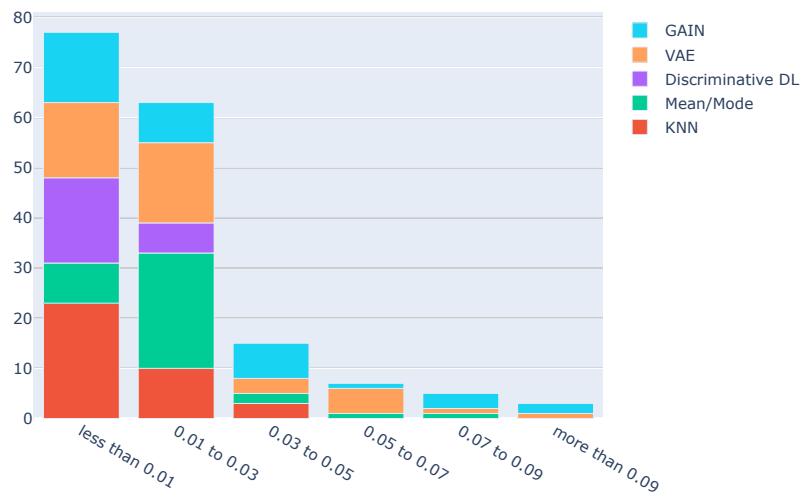


Abbildung B.6.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method

Improvement Relative to Average best for all Dataconstellations and best Imp

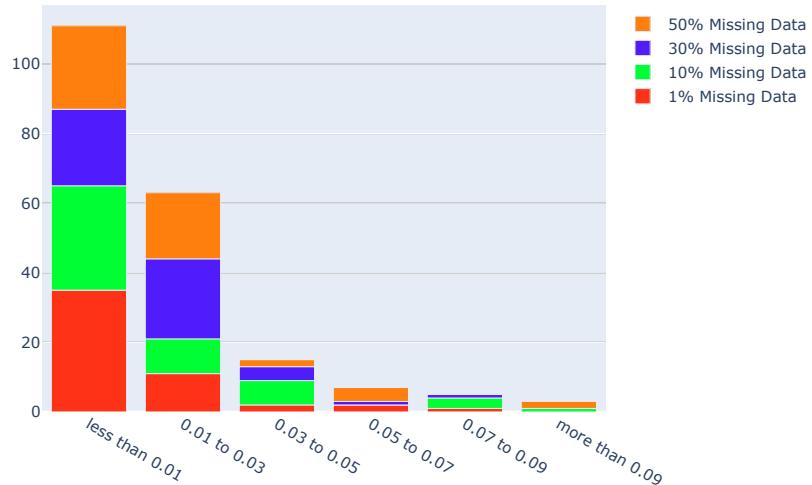


Abbildung B.7.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction

Improvement Relative to Average best for all Dataconstellations and best Imp

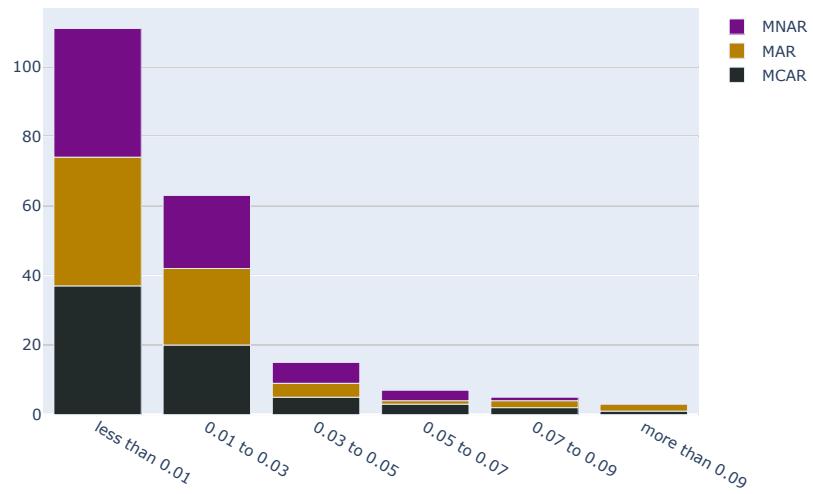


Abbildung B.8.: Improvement for Best Imputation Method per Experiment Scenario
Relative to the Average Best Imputation Method by F1 Score Points
by Missingness Pattern

MAR - 0.01

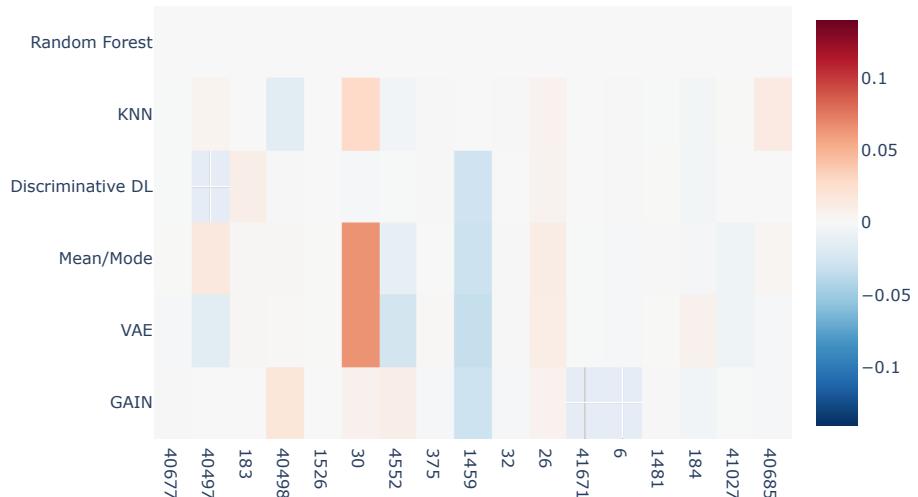


Abbildung B.9.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction

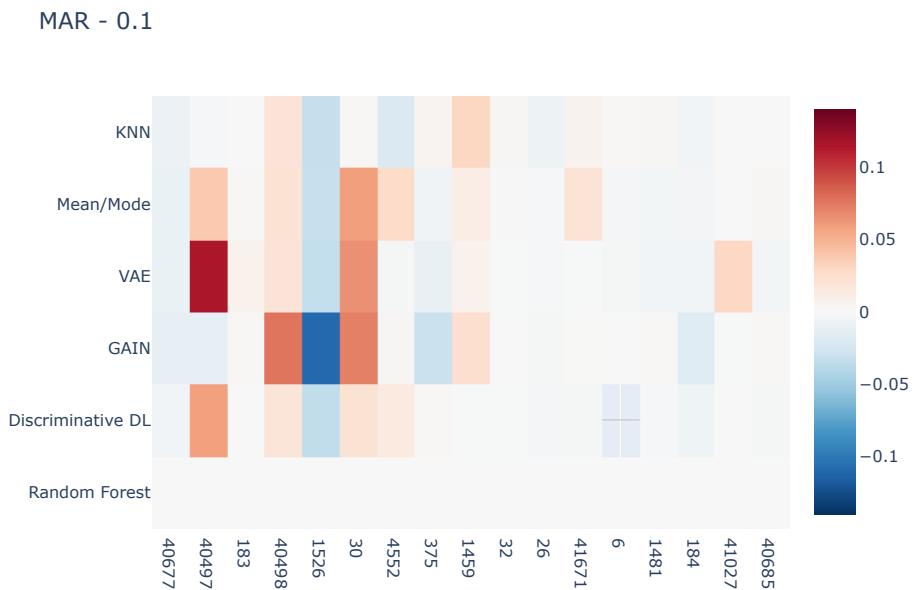


Abbildung B.10.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction

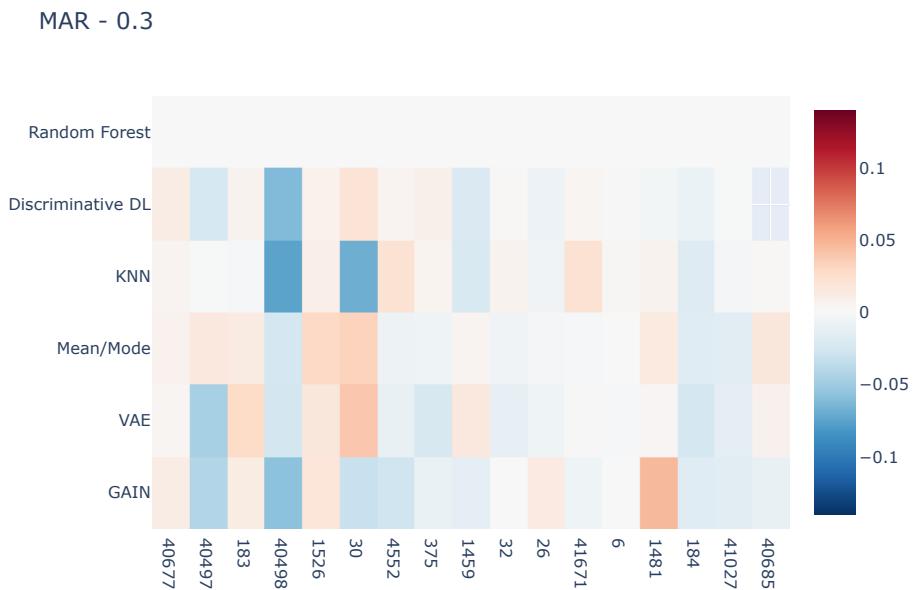


Abbildung B.11.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction

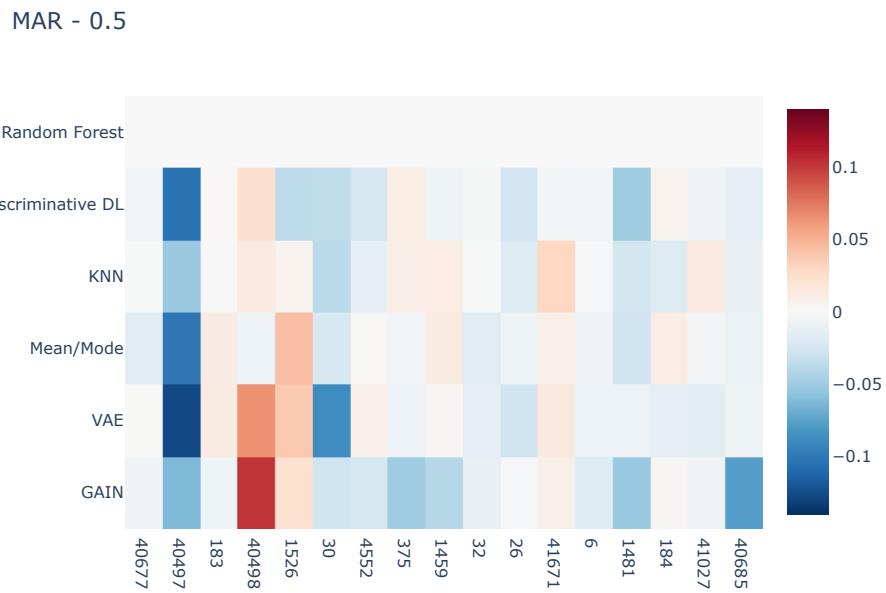


Abbildung B.12.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction

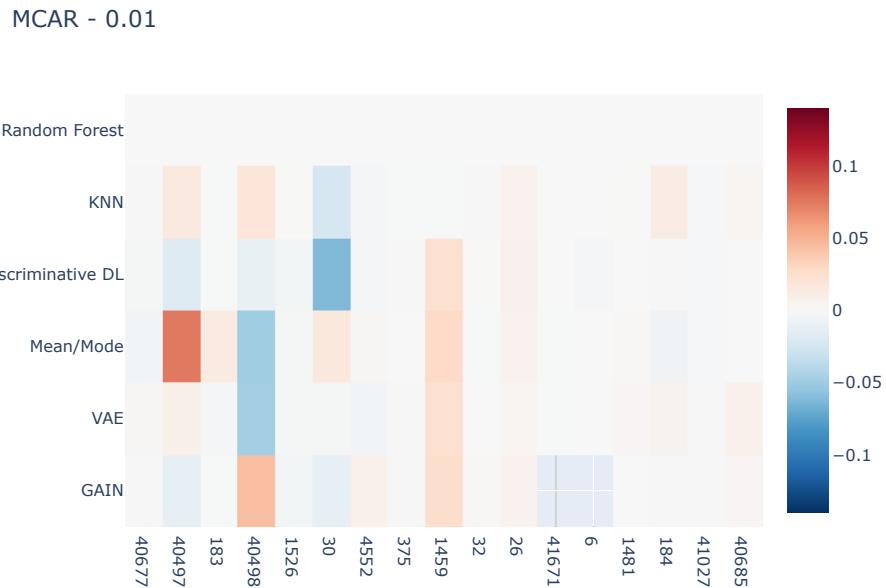


Abbildung B.13.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction

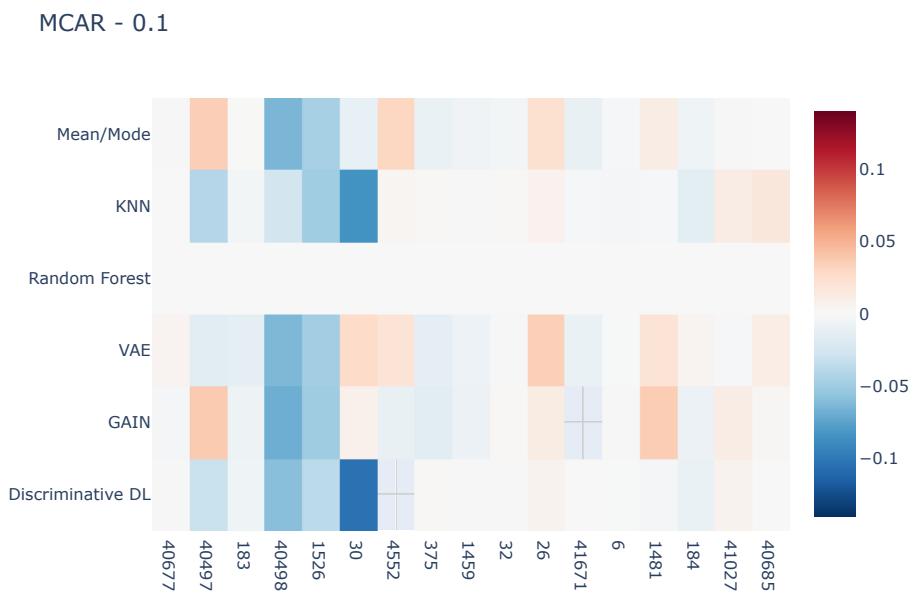


Abbildung B.14.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction

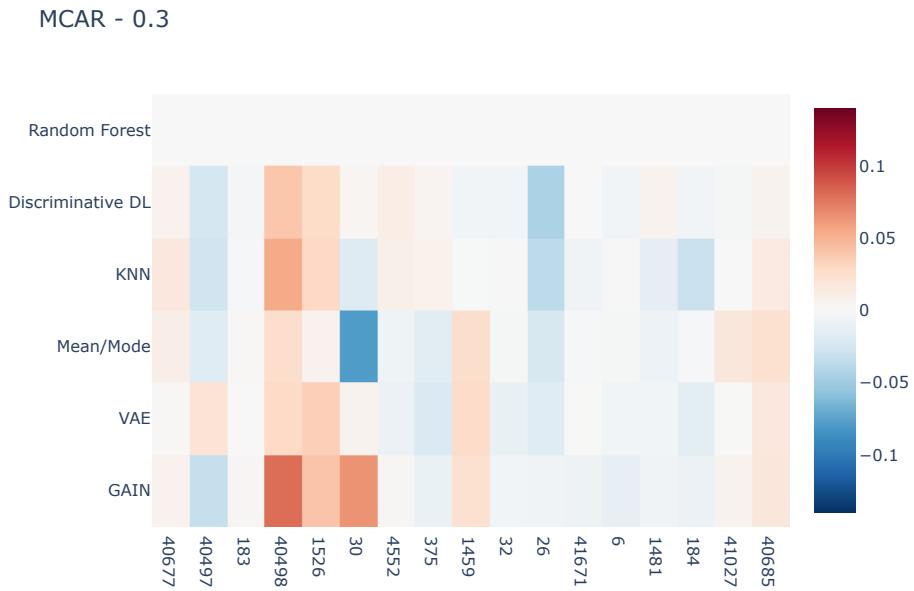


Abbildung B.15.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction

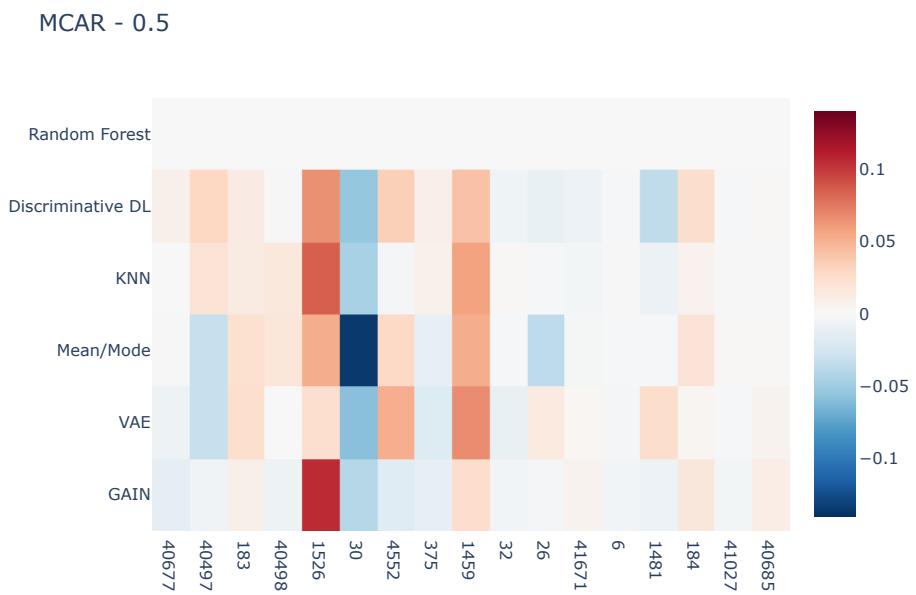


Abbildung B.16.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction

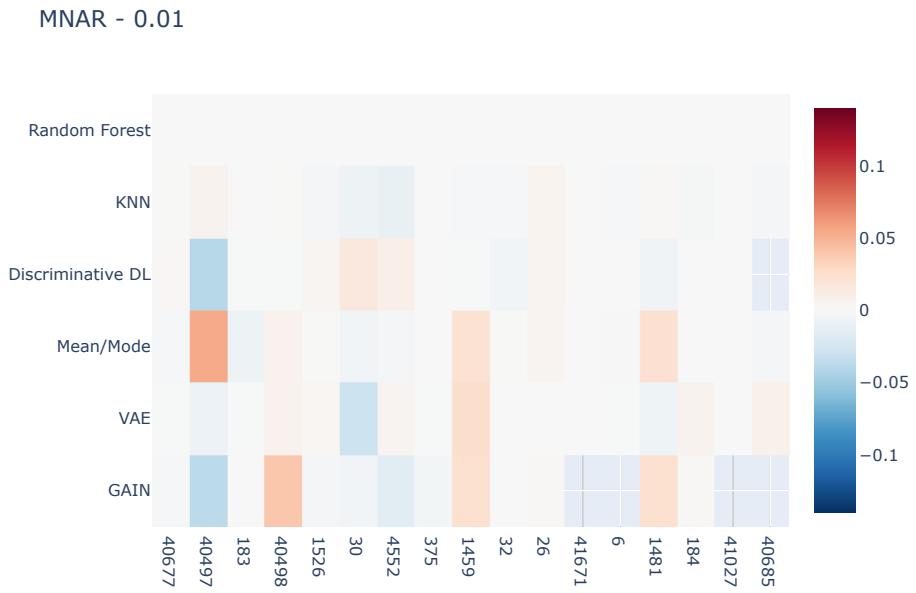


Abbildung B.17.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction

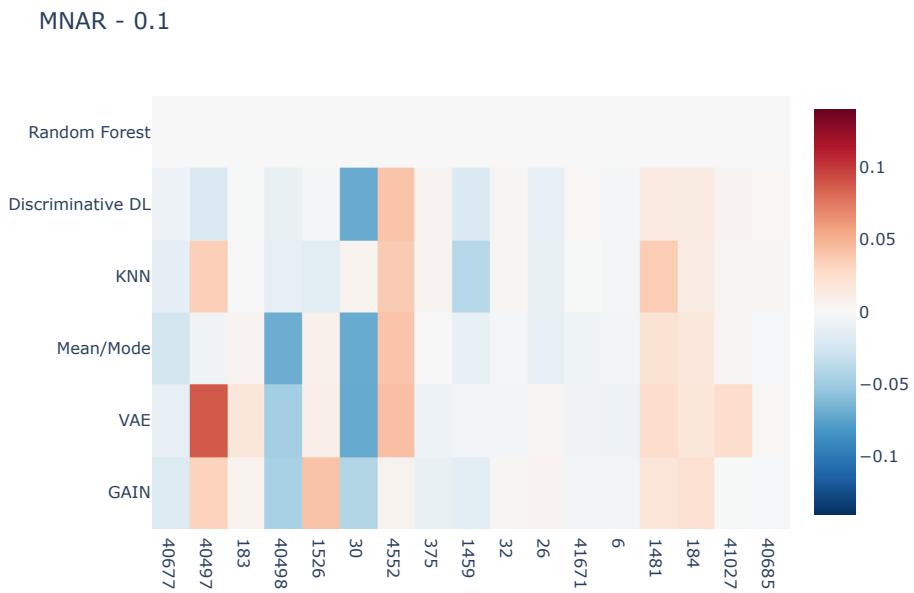


Abbildung B.18.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction

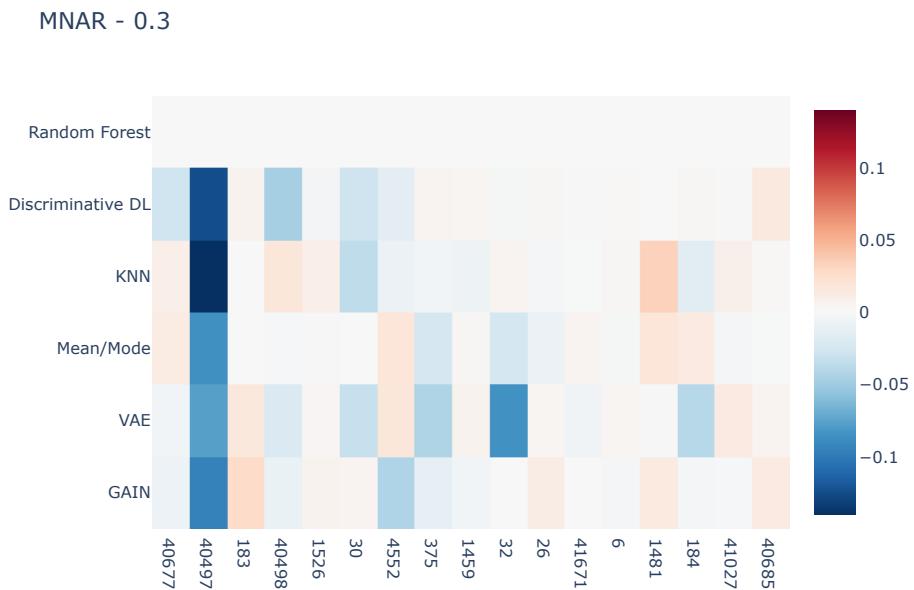


Abbildung B.19.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction

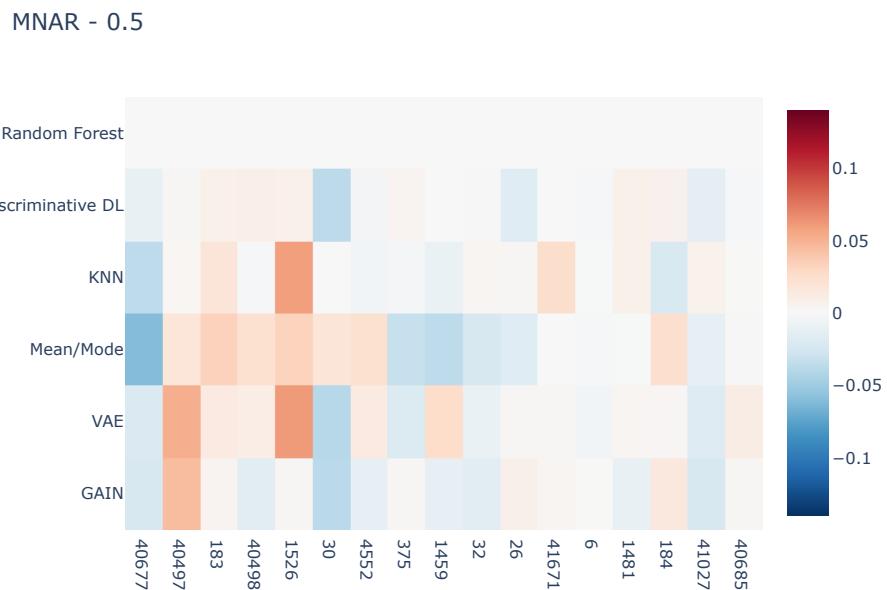


Abbildung B.20.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction

C. Appendix Regression Experiments

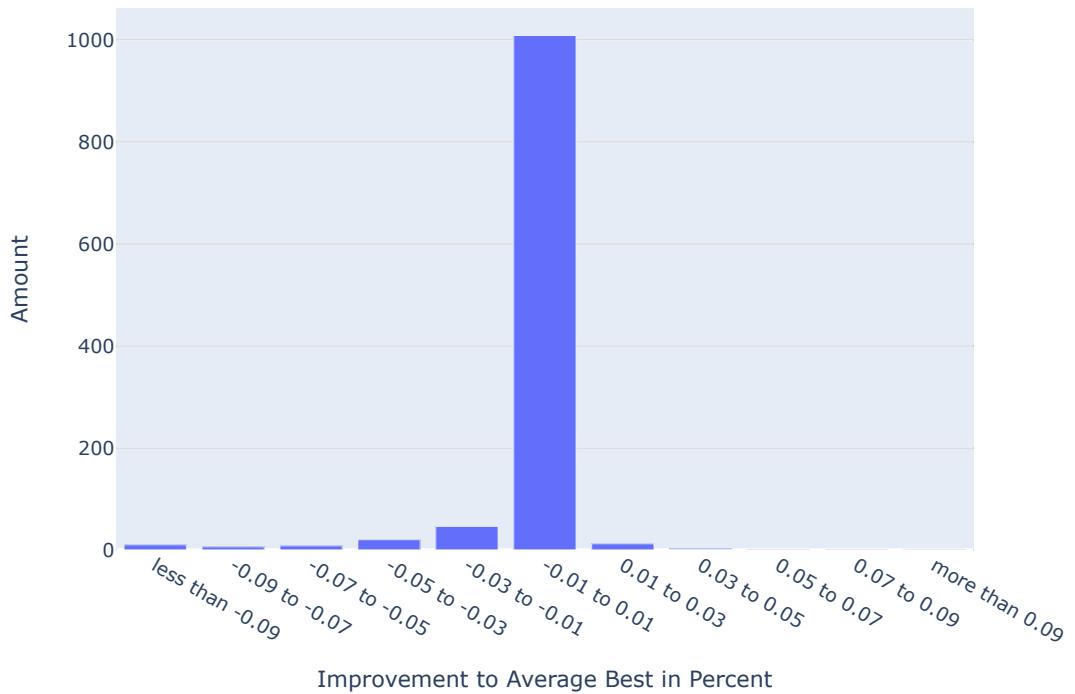


Abbildung C.1.: Improvement for Experiments Relative to the Average Best Imputation Method by Percentage

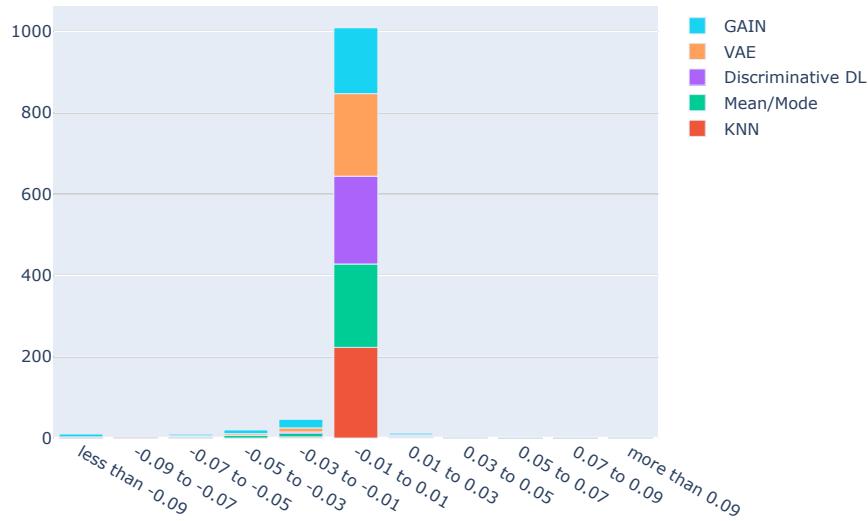


Abbildung C.2.: Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Imputation Method

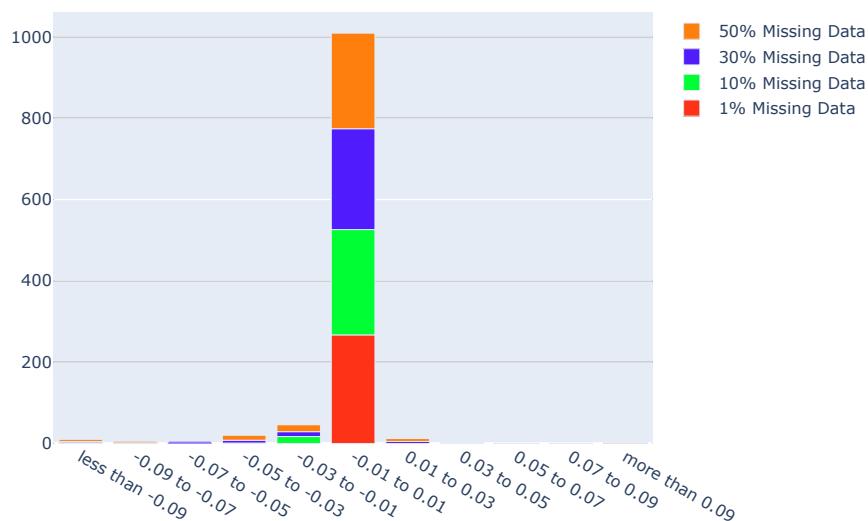


Abbildung C.3.: Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Missing Fraction

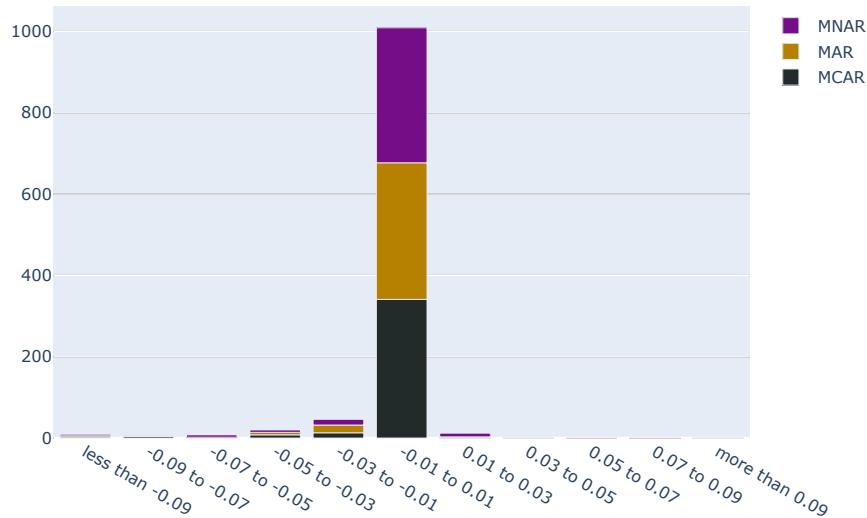


Abbildung C.4.: Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Missingness Pattern

Improvement Relative to Average best the best Imputation Method for each D

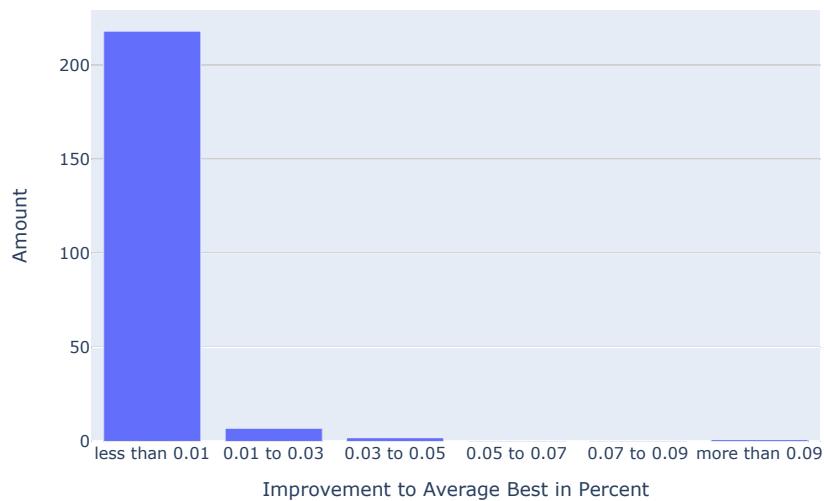


Abbildung C.5.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage

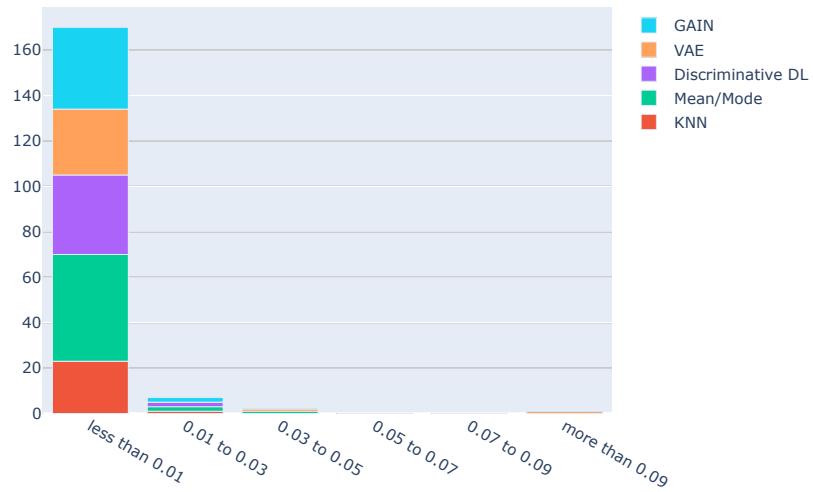


Abbildung C.6.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Imputation Method

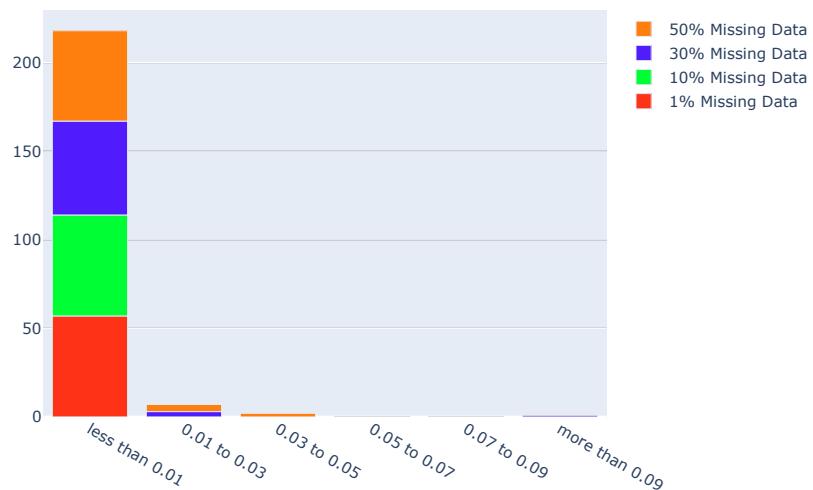


Abbildung C.7.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Missing Fraction

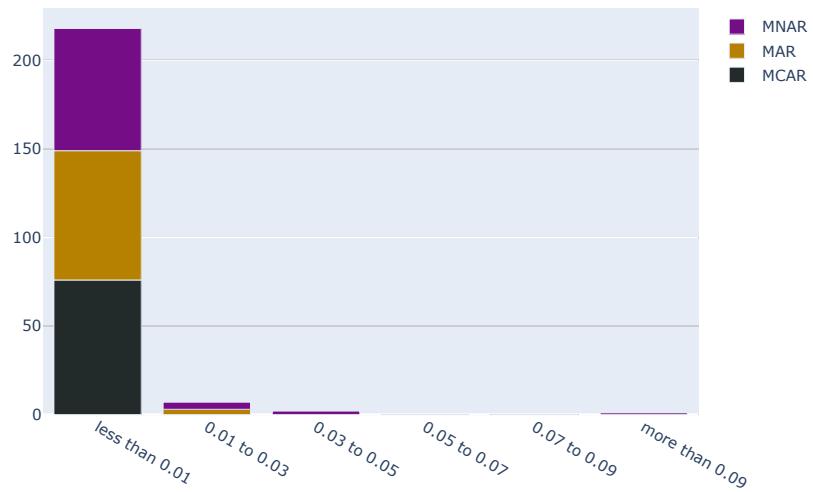


Abbildung C.8.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Missingness Pattern

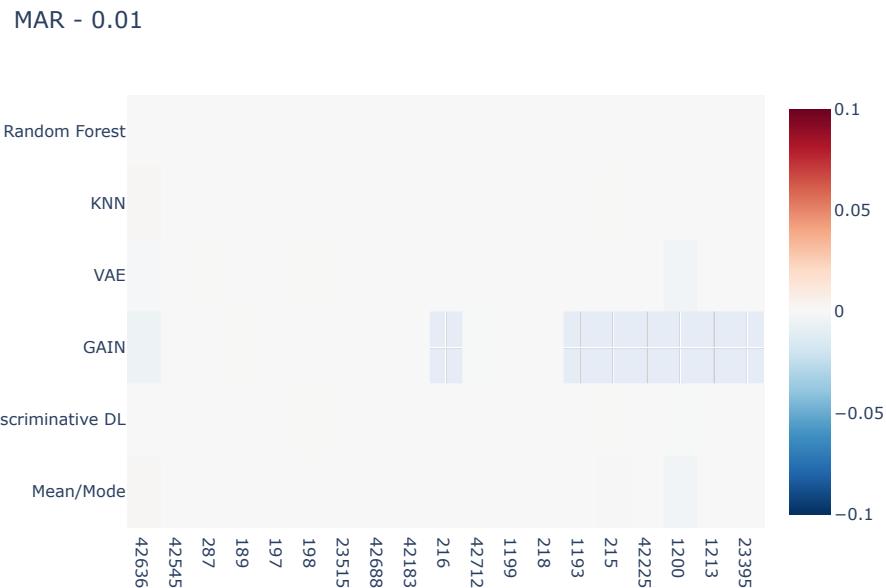


Abbildung C.9.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction

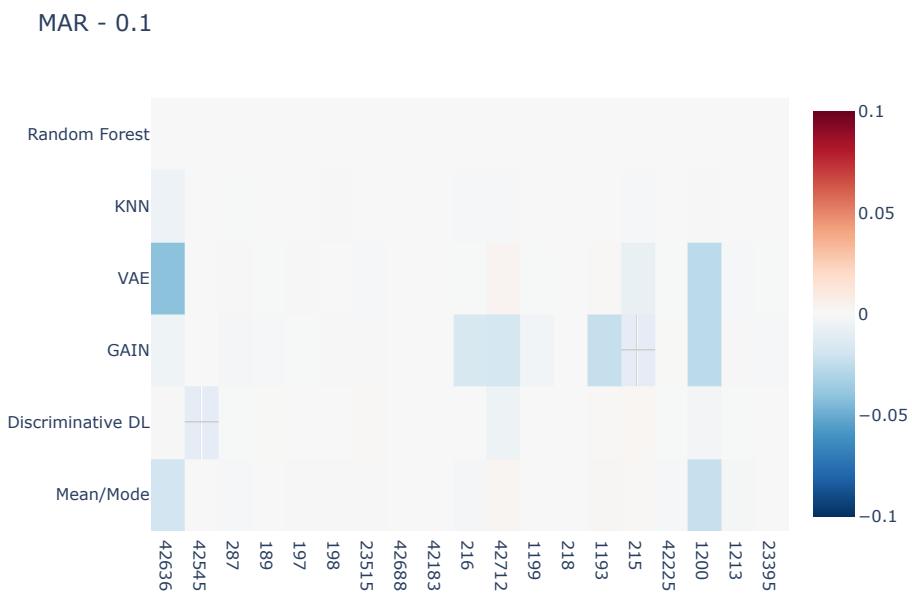


Abbildung C.10.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction

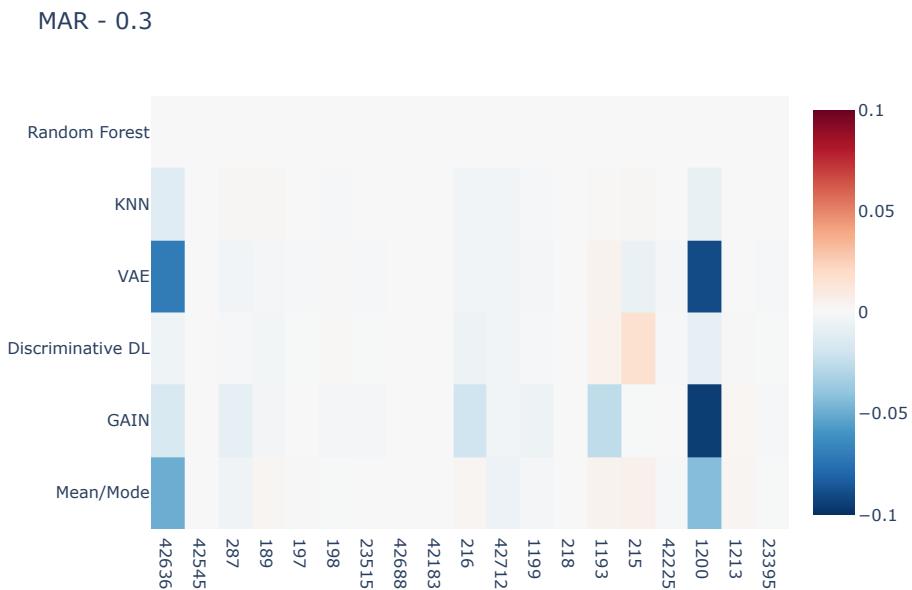


Abbildung C.11.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction

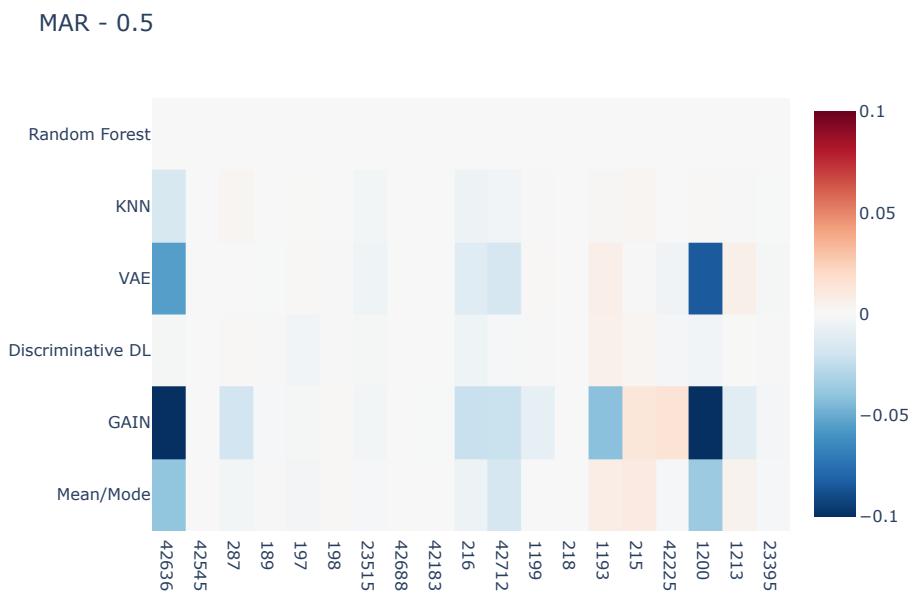


Abbildung C.12.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction

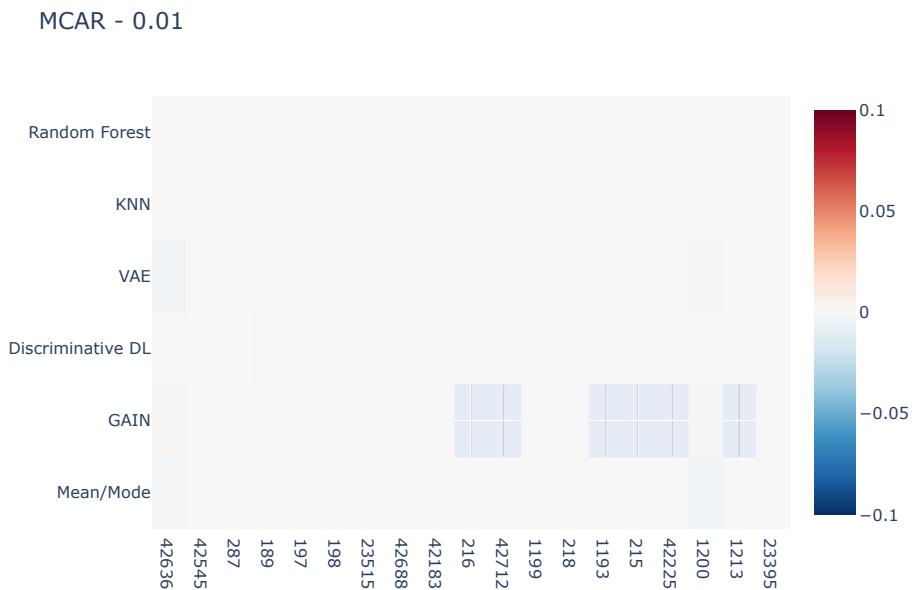


Abbildung C.13.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction

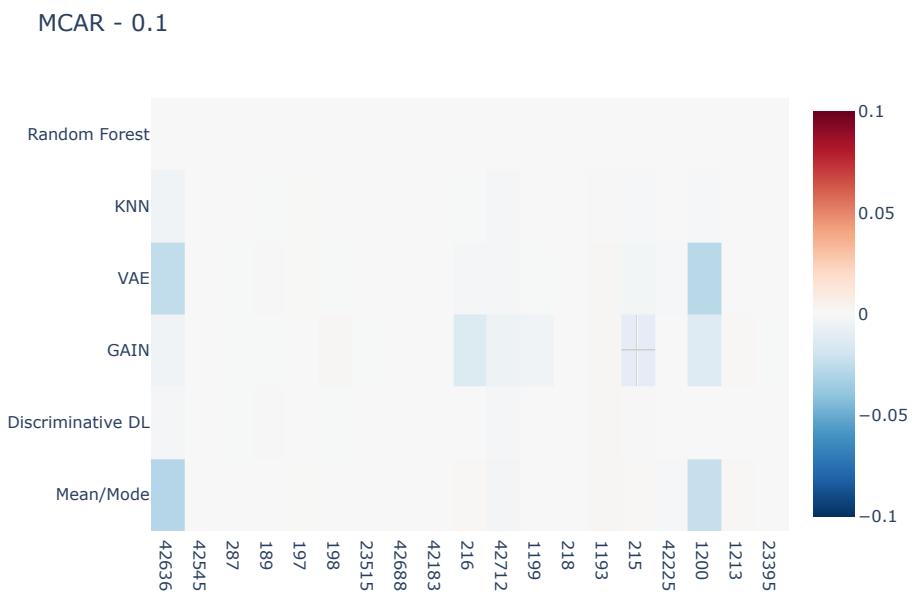


Abbildung C.14.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction

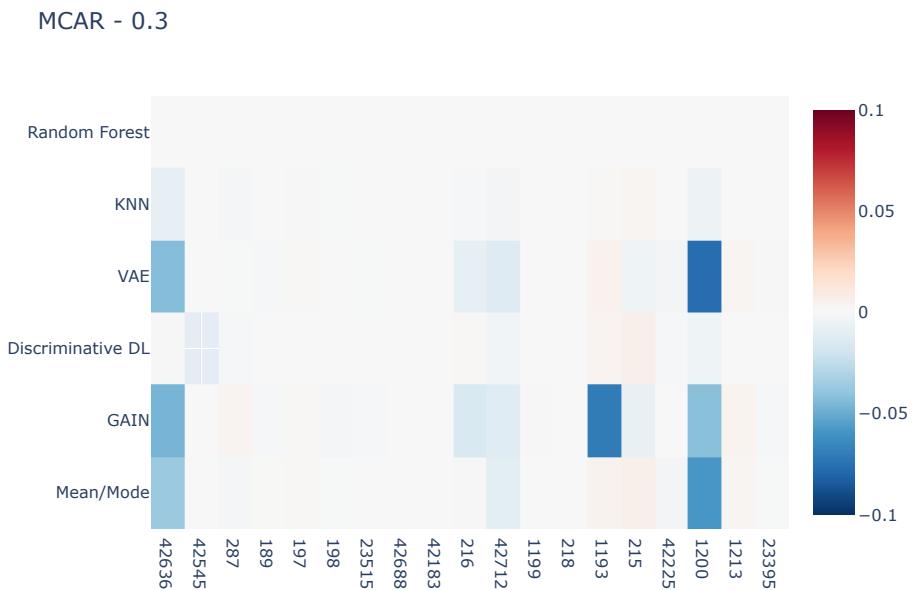


Abbildung C.15.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction

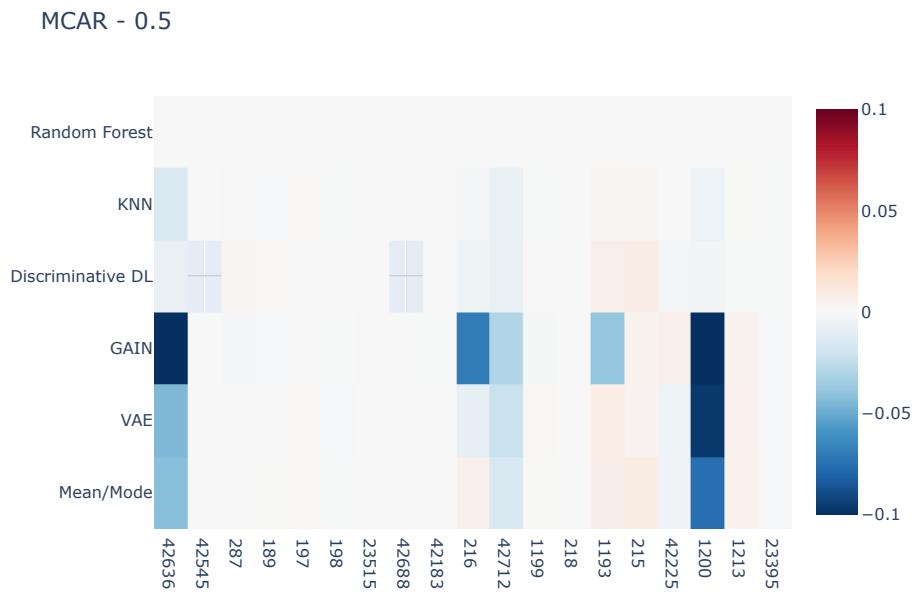


Abbildung C.16.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction

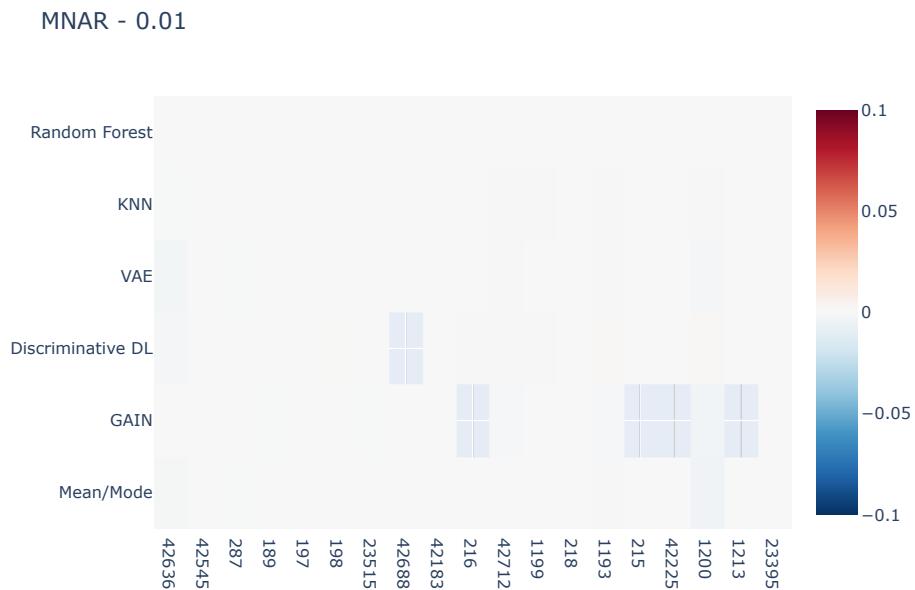


Abbildung C.17.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction

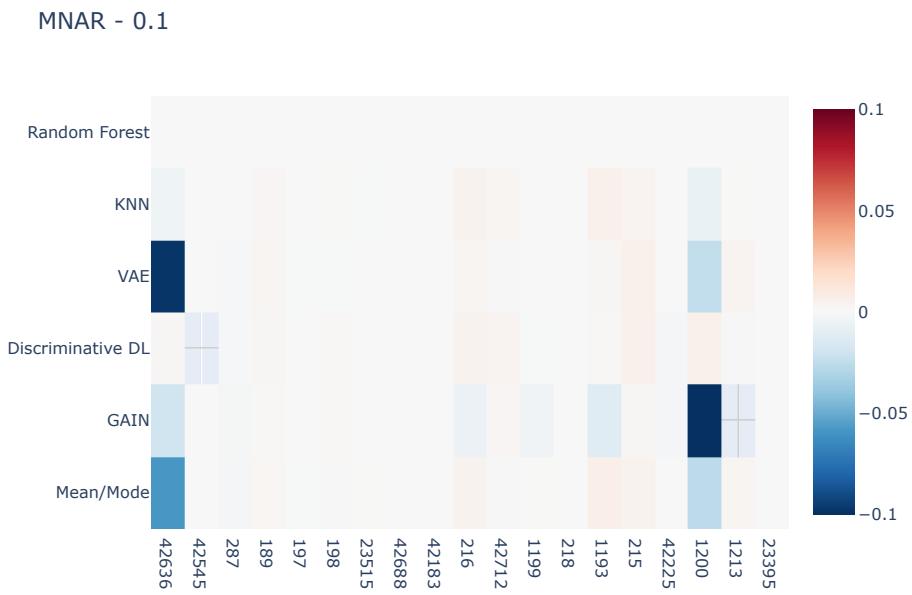


Abbildung C.18.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction

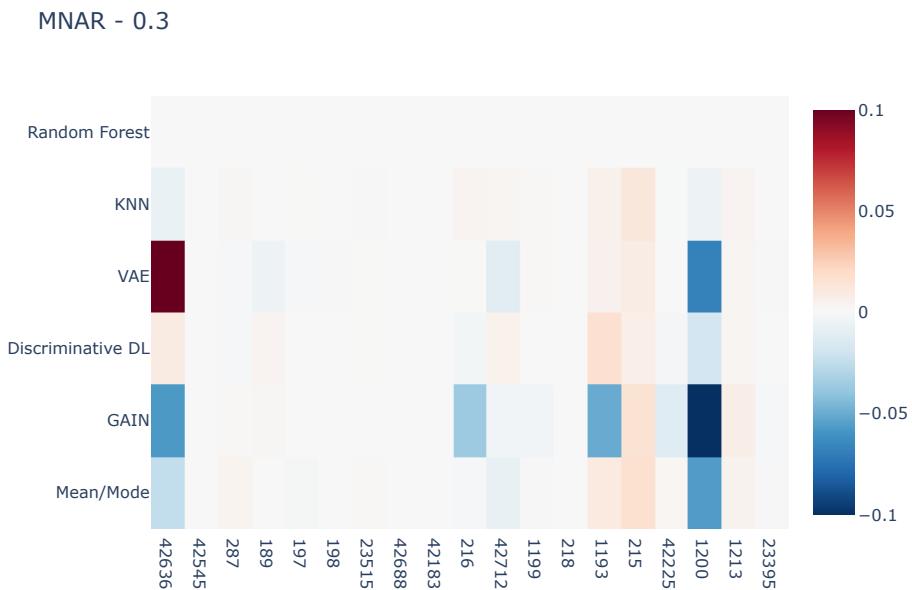


Abbildung C.19.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction

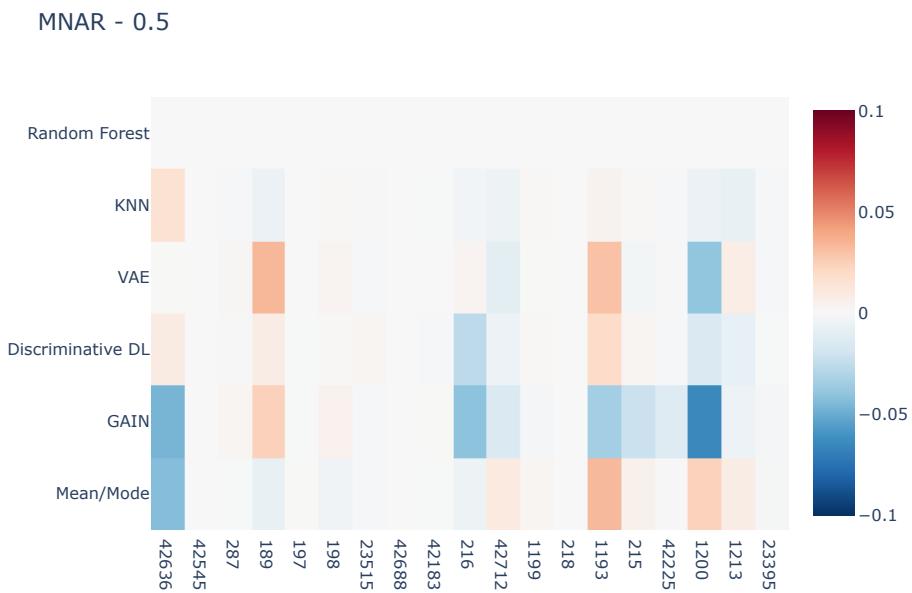


Abbildung C.20.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction

D. Appendix Binary Classification Subset Experiments

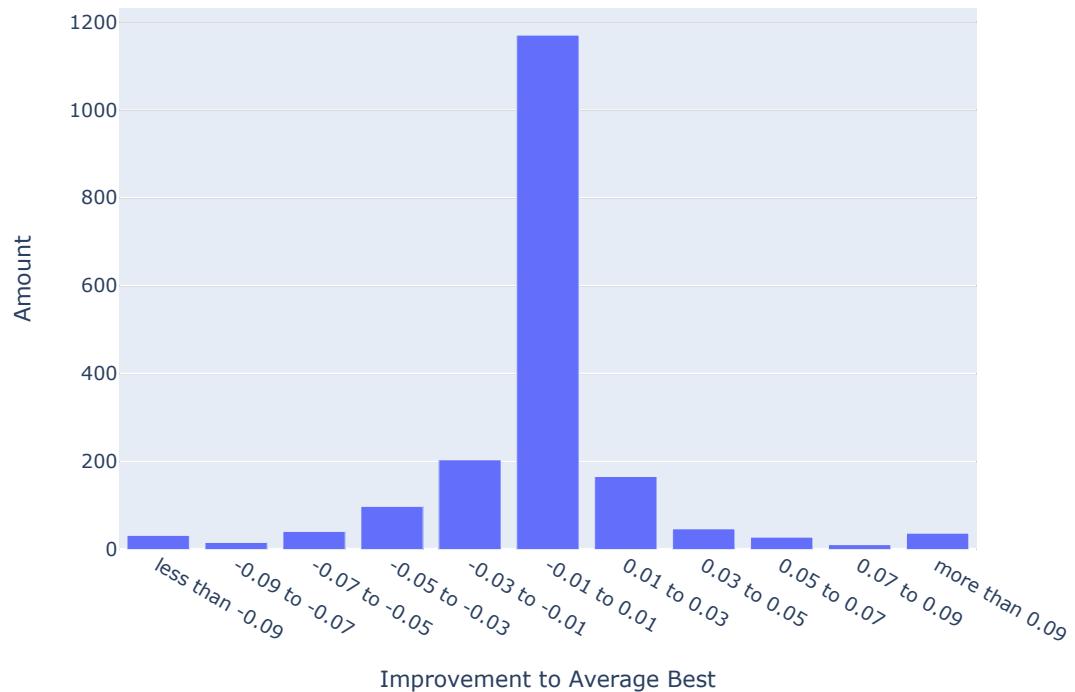


Abbildung D.1.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points

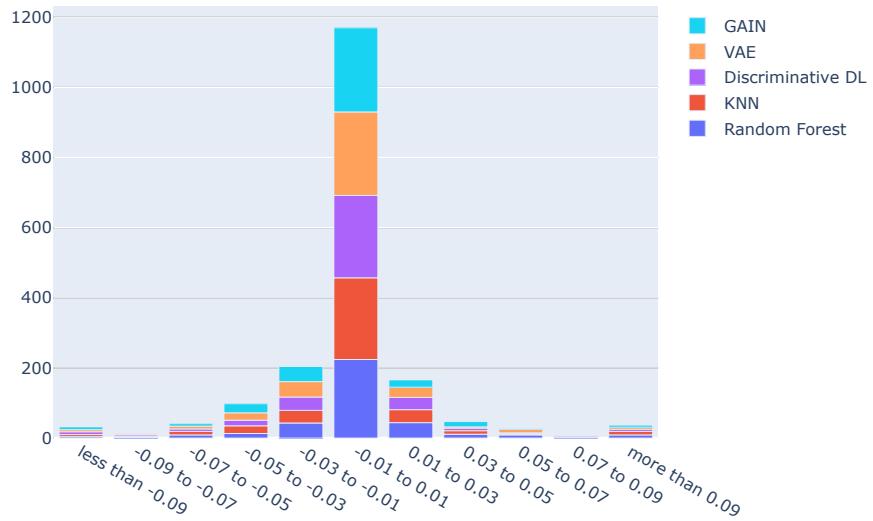


Abbildung D.2.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Imputation Methods

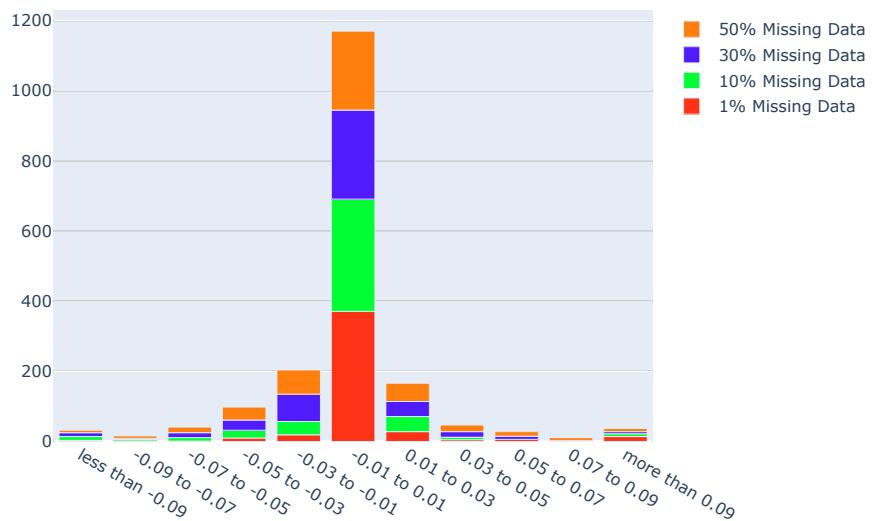


Abbildung D.3.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction

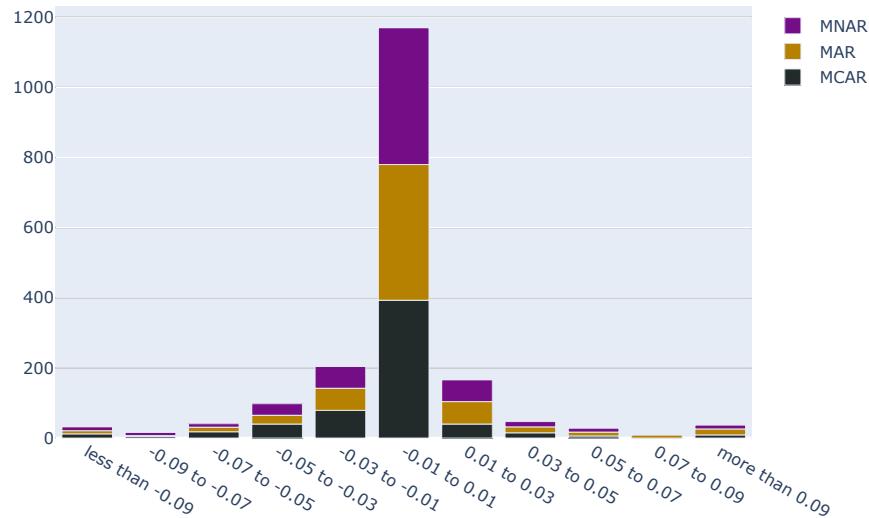


Abbildung D.4.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missingness Pattern

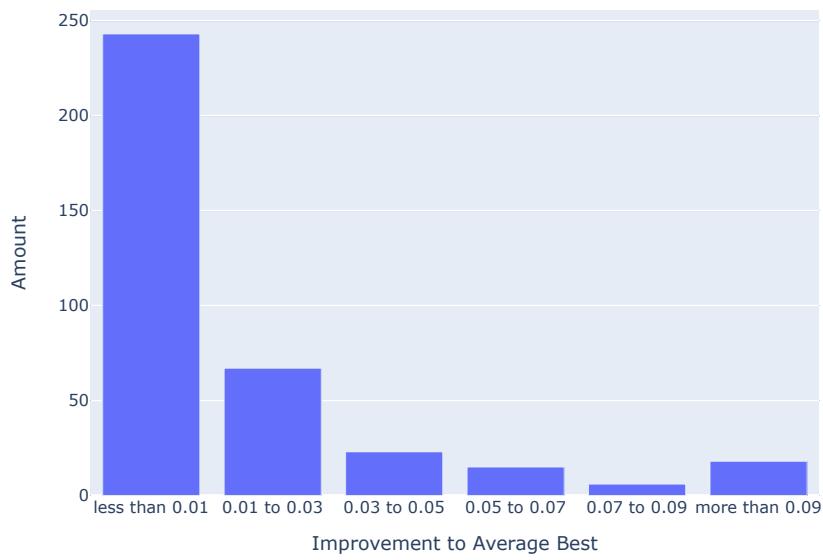


Abbildung D.5.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points

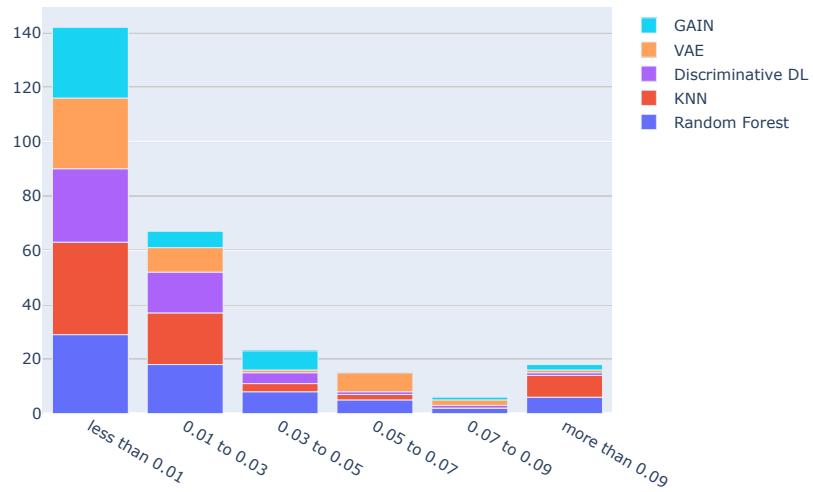


Abbildung D.6.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method

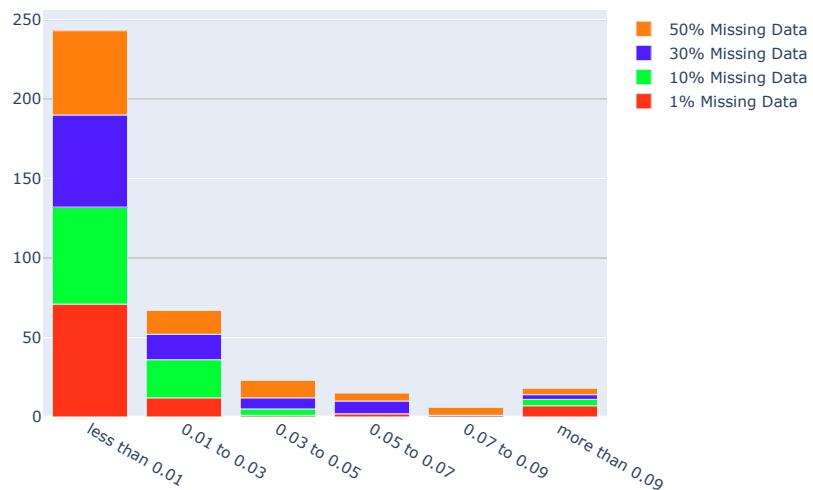


Abbildung D.7.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction

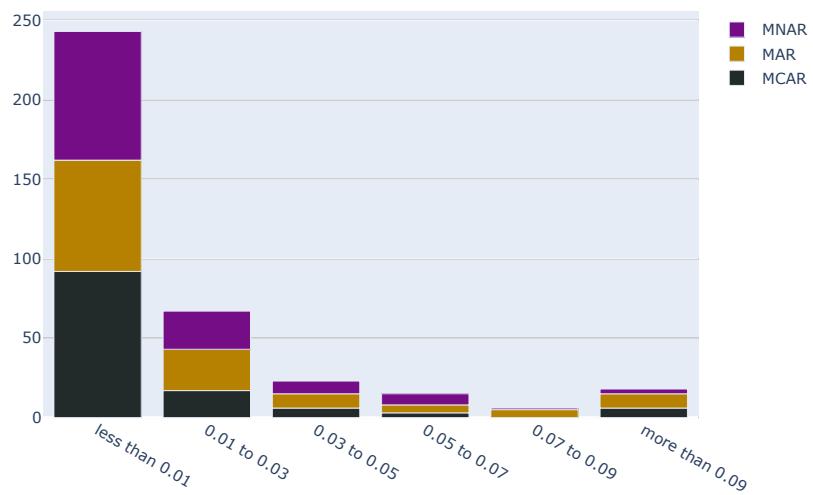


Abbildung D.8.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points Missingness Pattern

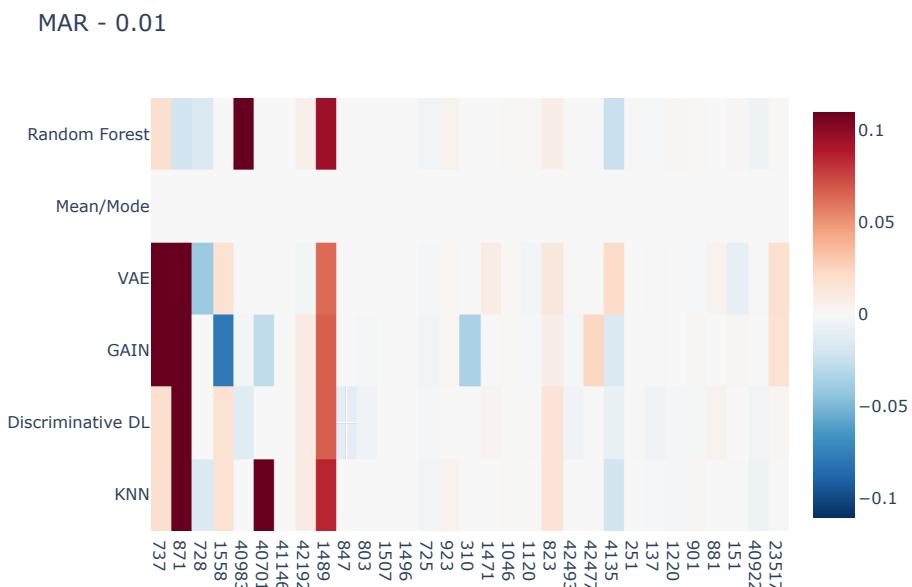


Abbildung D.9.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction

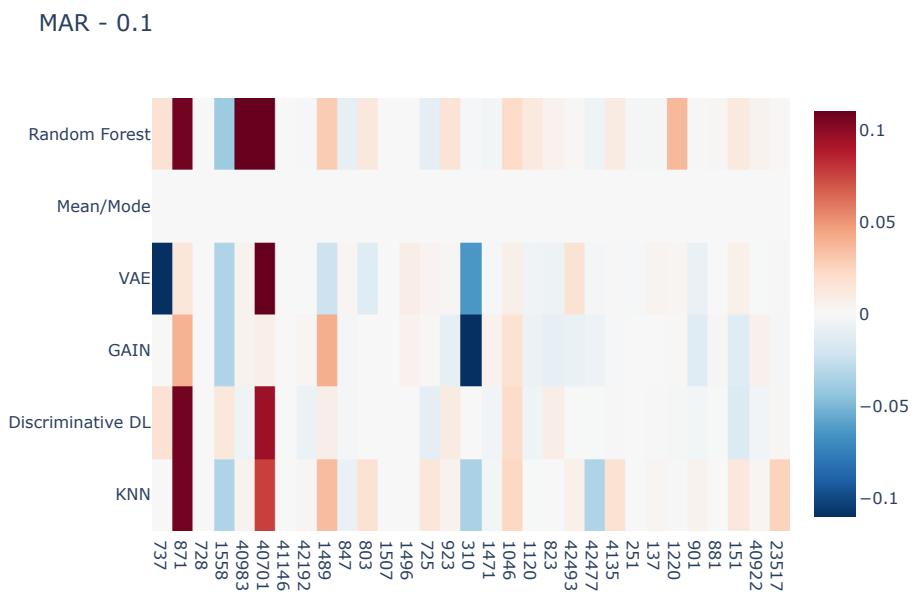


Abbildung D.10.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction

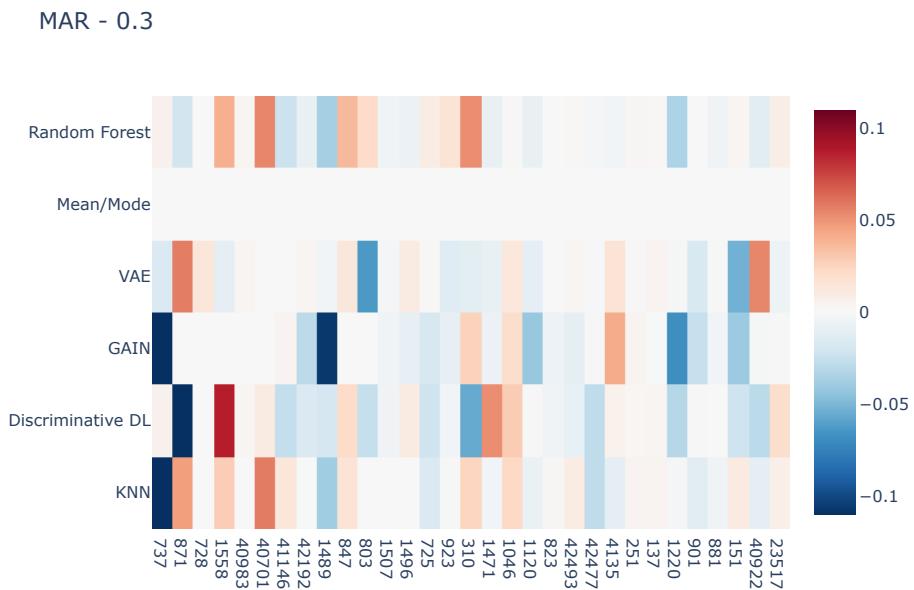


Abbildung D.11.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction

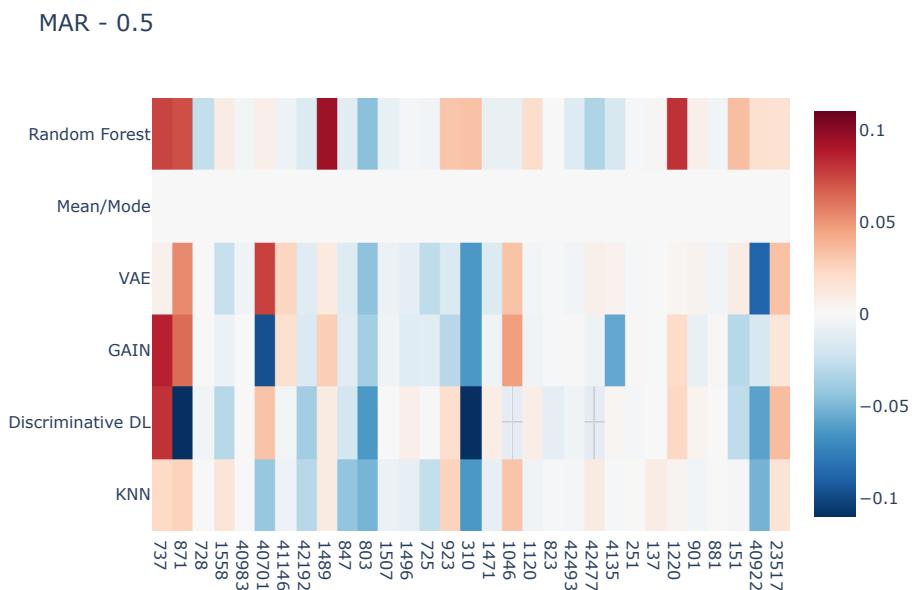


Abbildung D.12.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction

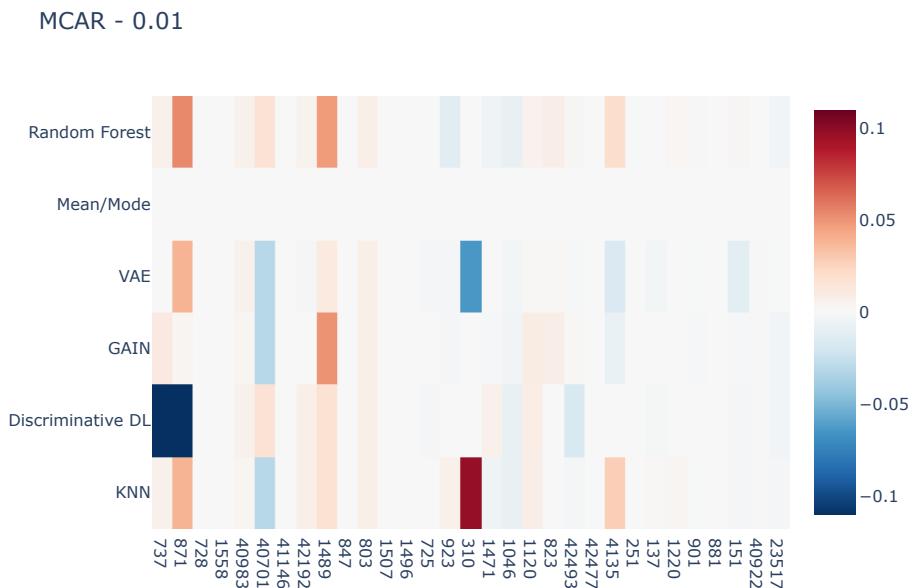


Abbildung D.13.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction

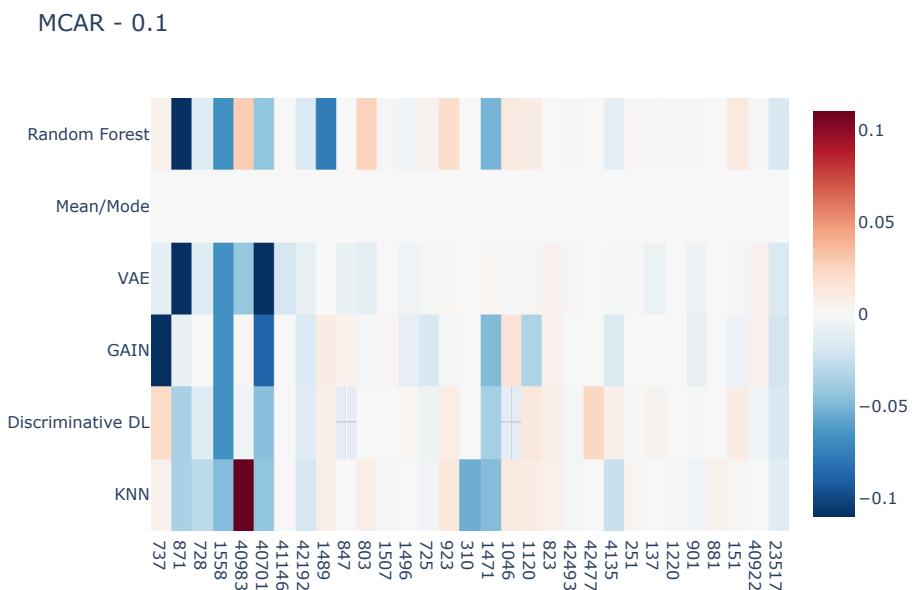


Abbildung D.14.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction

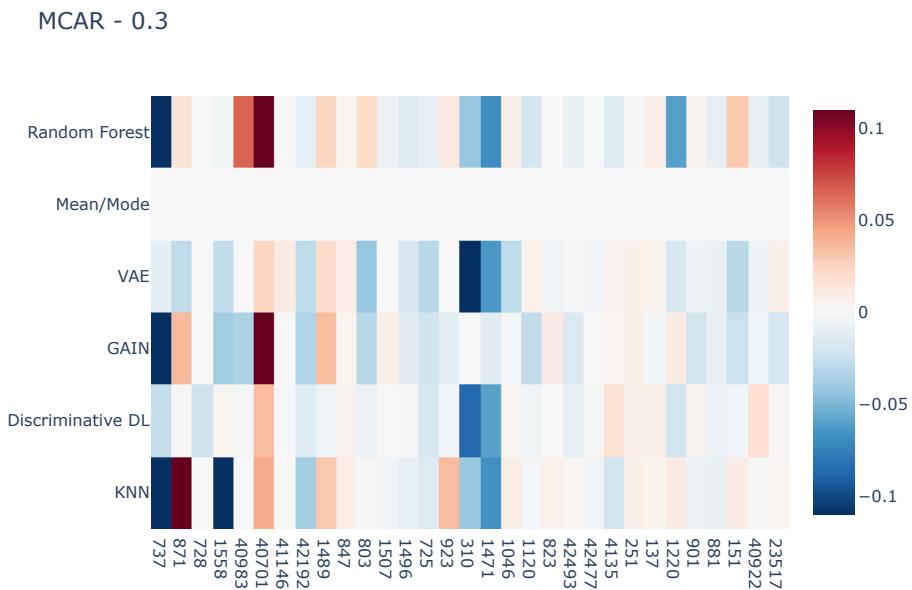


Abbildung D.15.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction

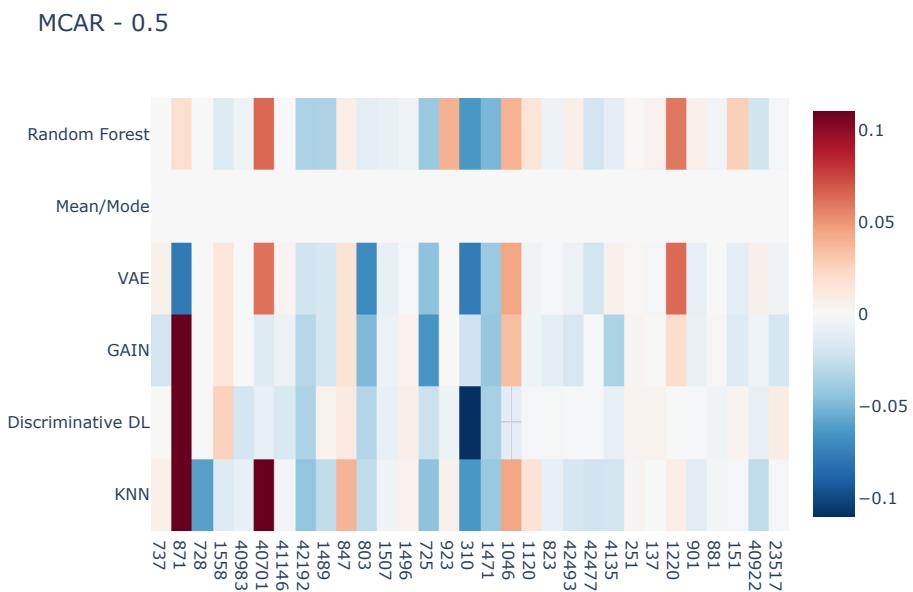


Abbildung D.16.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction

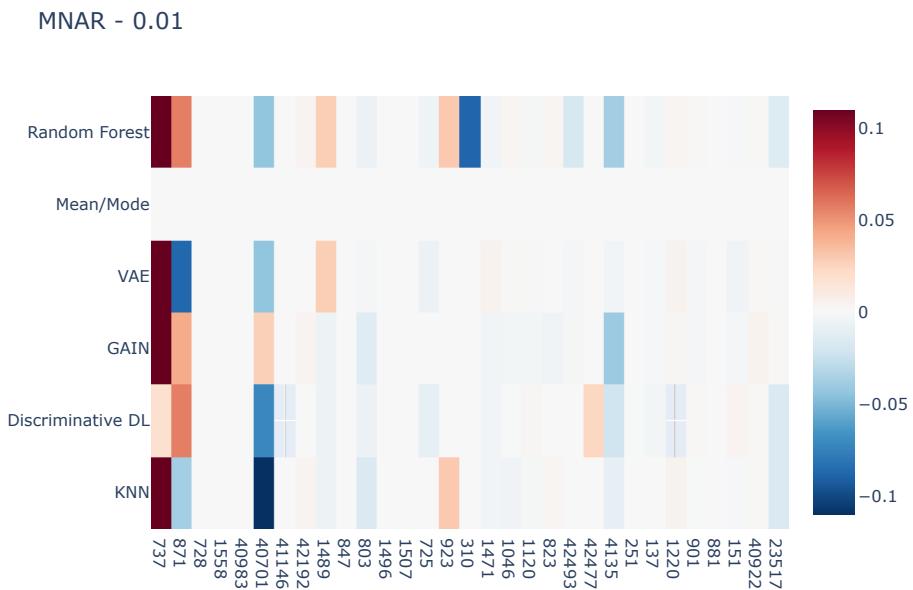


Abbildung D.17.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction

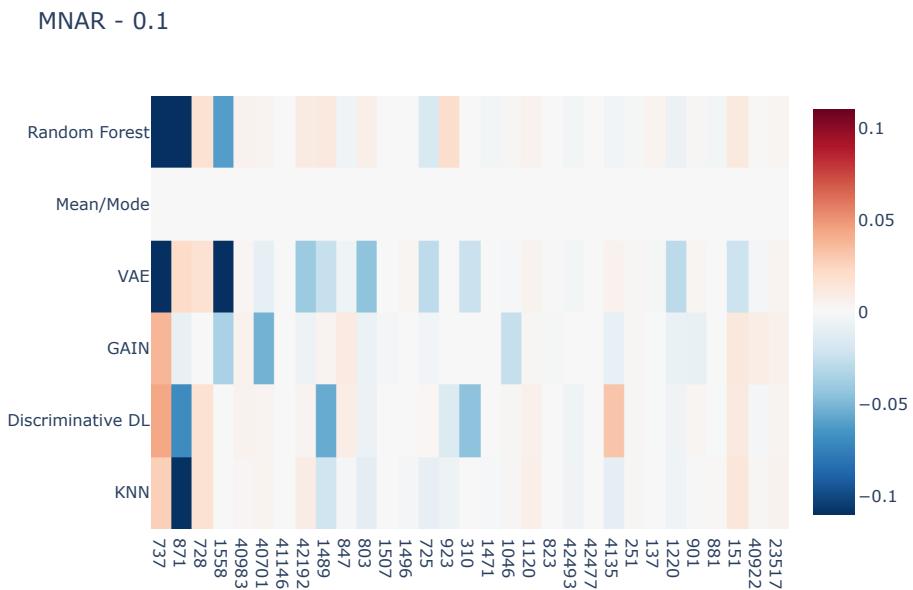


Abbildung D.18.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction

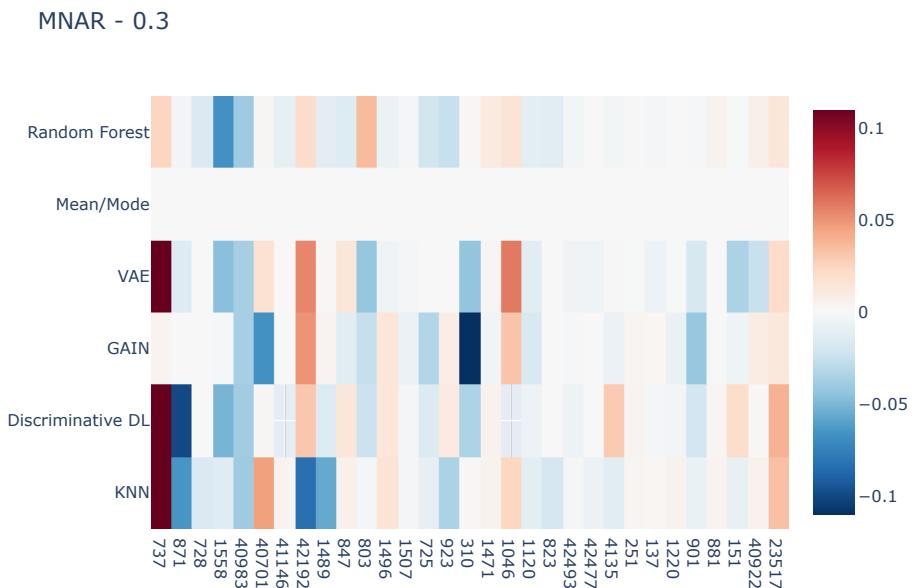


Abbildung D.19.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction

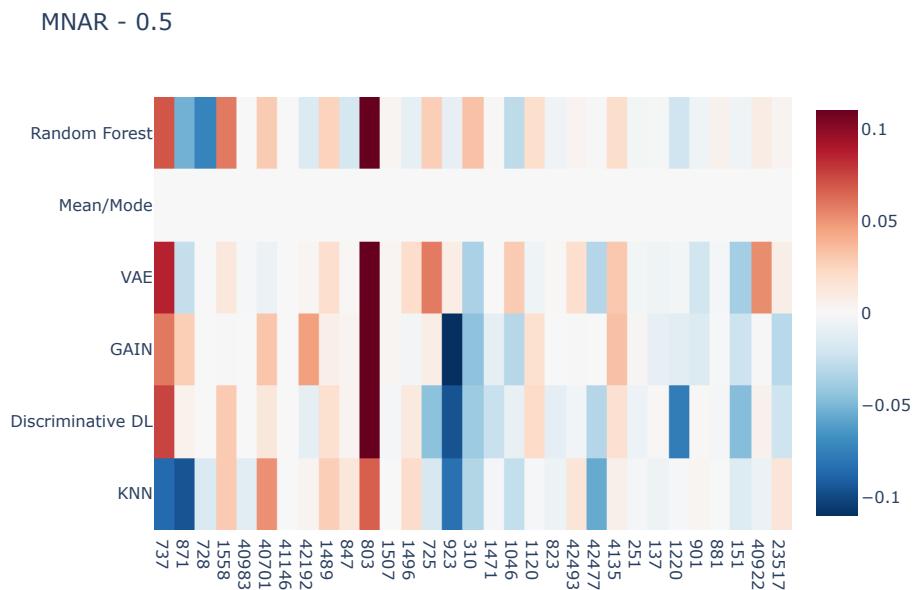


Abbildung D.20.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction

E. Appendix Multiclass Classification Subset Experiments

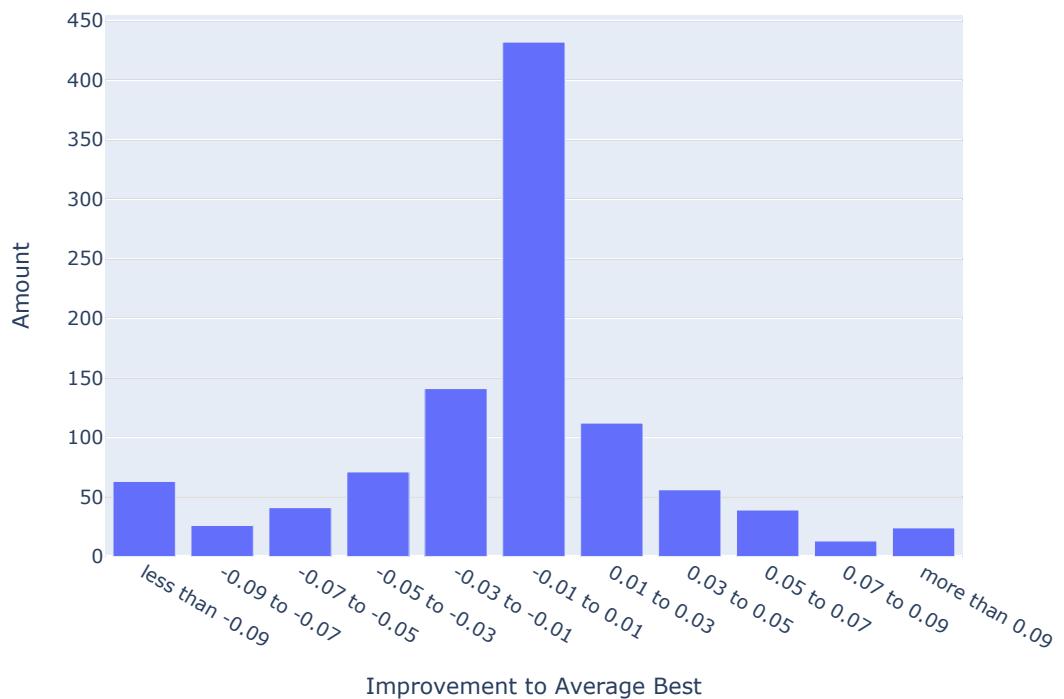


Abbildung E.1.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points

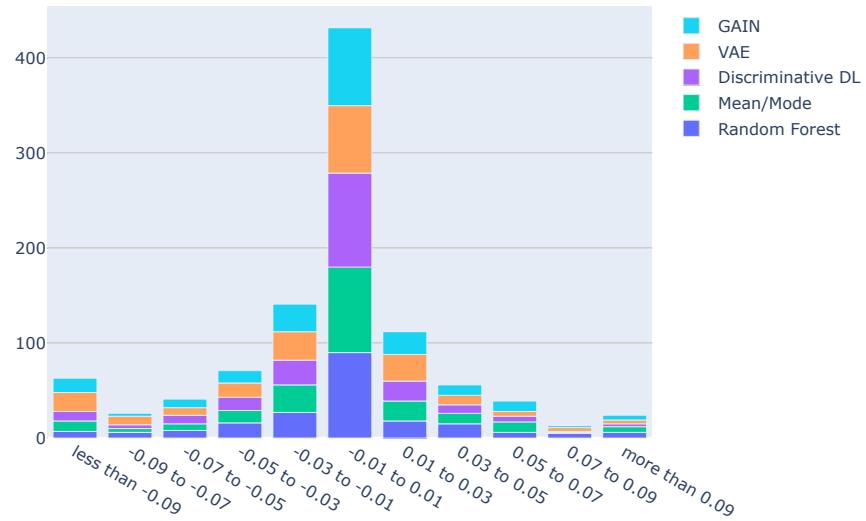


Abbildung E.2.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method

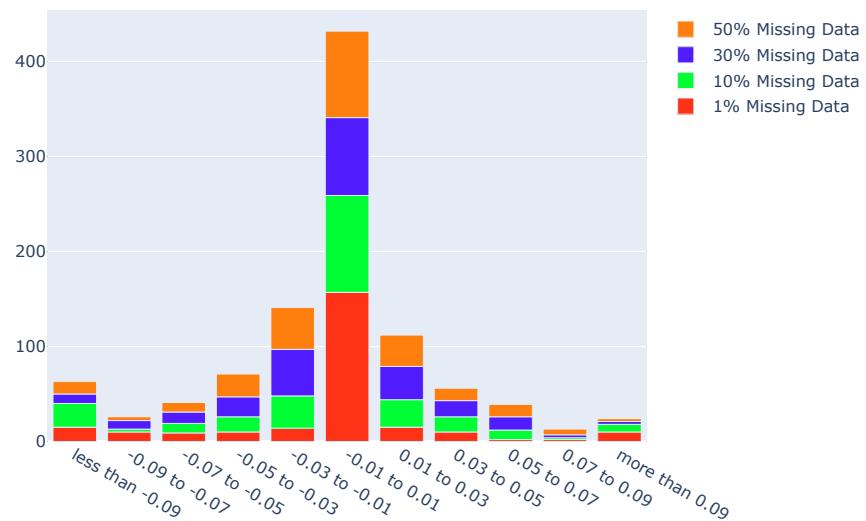


Abbildung E.3.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction

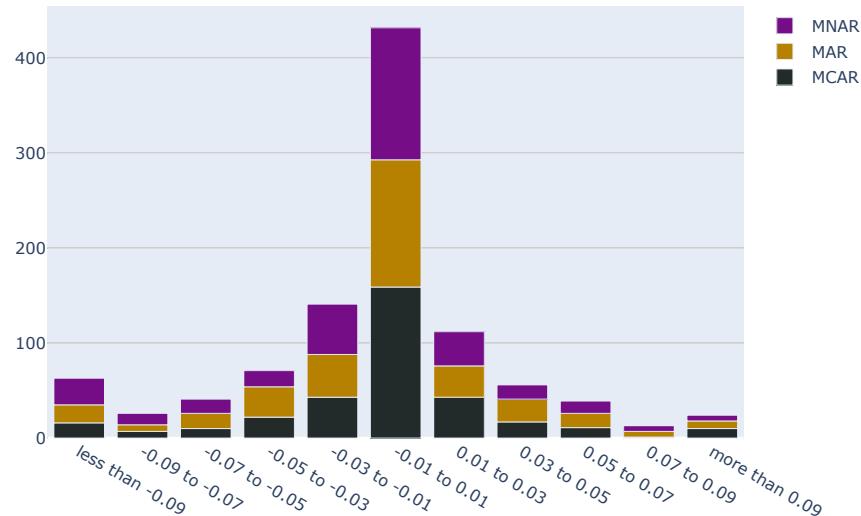


Abbildung E.4.: Improvement for Experiments Relative to the Average Best Imputation Method by F1 Score Points by Missingness Pattern

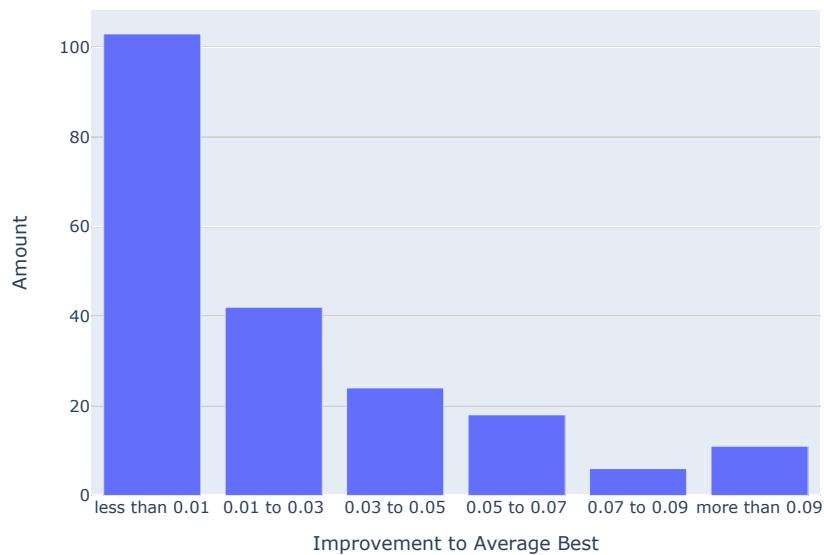


Abbildung E.5.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points

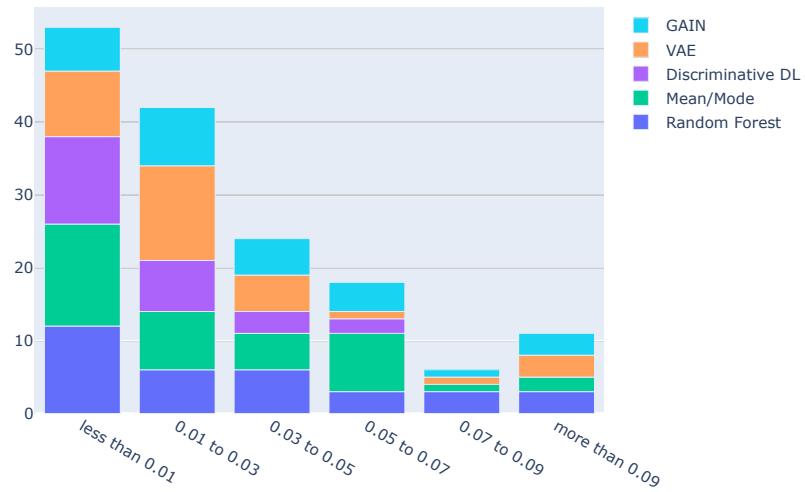


Abbildung E.6.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Imputation Method

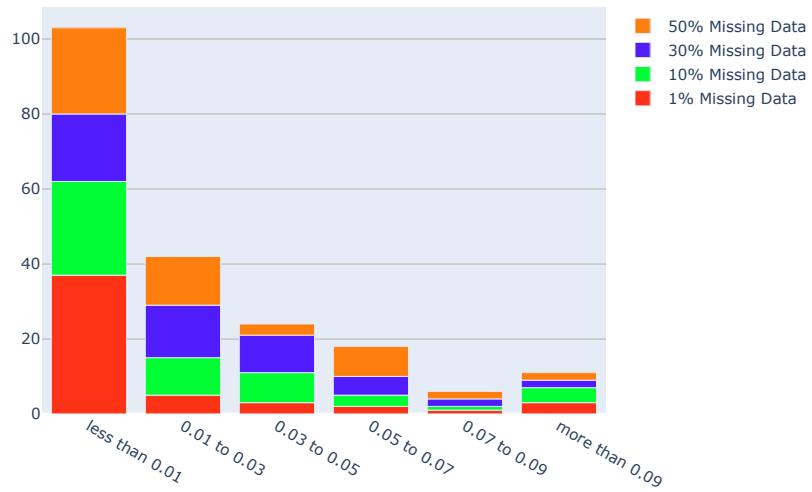


Abbildung E.7.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Missing Fraction

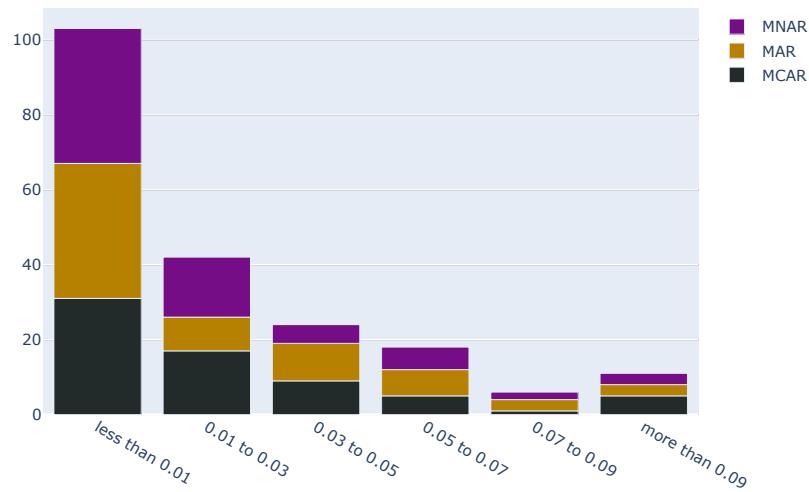


Abbildung E.8.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by F1 Score Points by Missingness Pattern

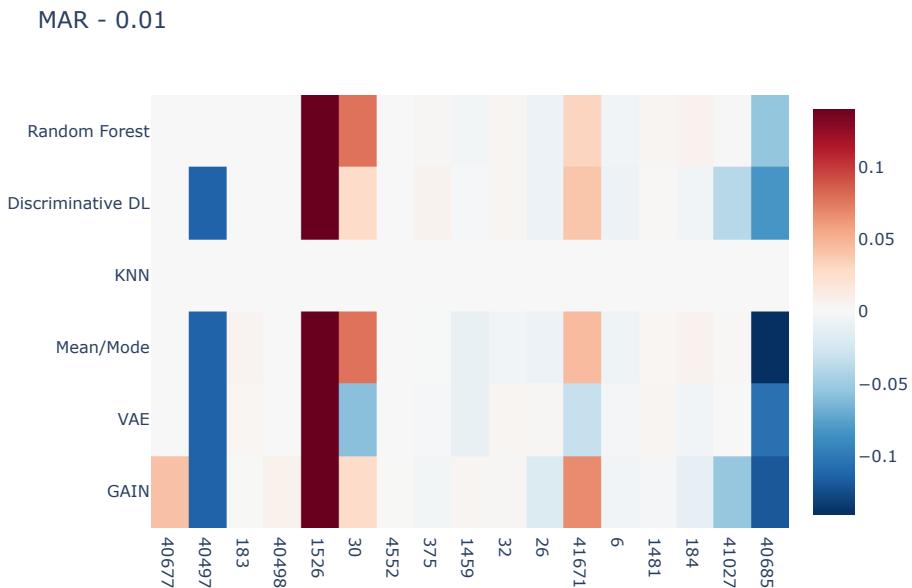


Abbildung E.9.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction

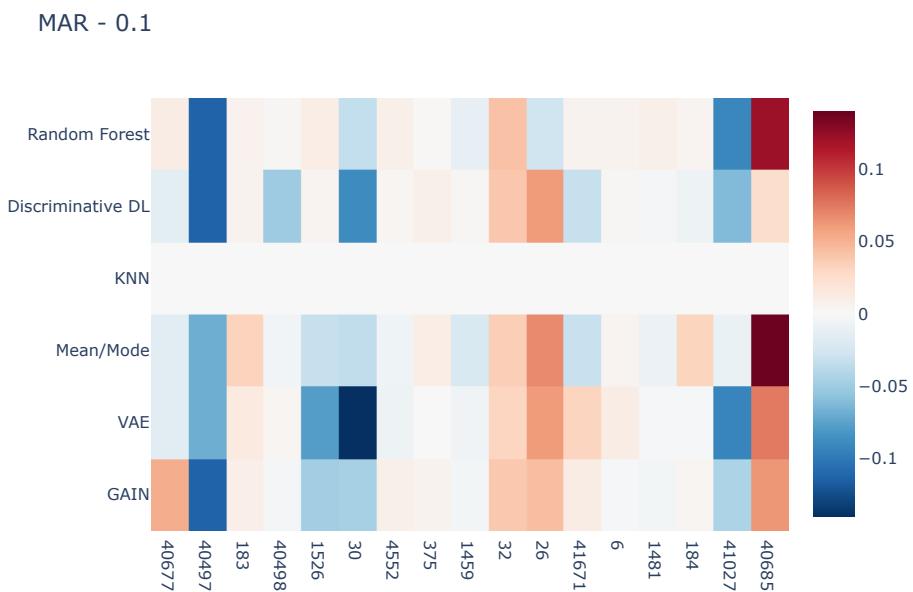


Abbildung E.10.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction

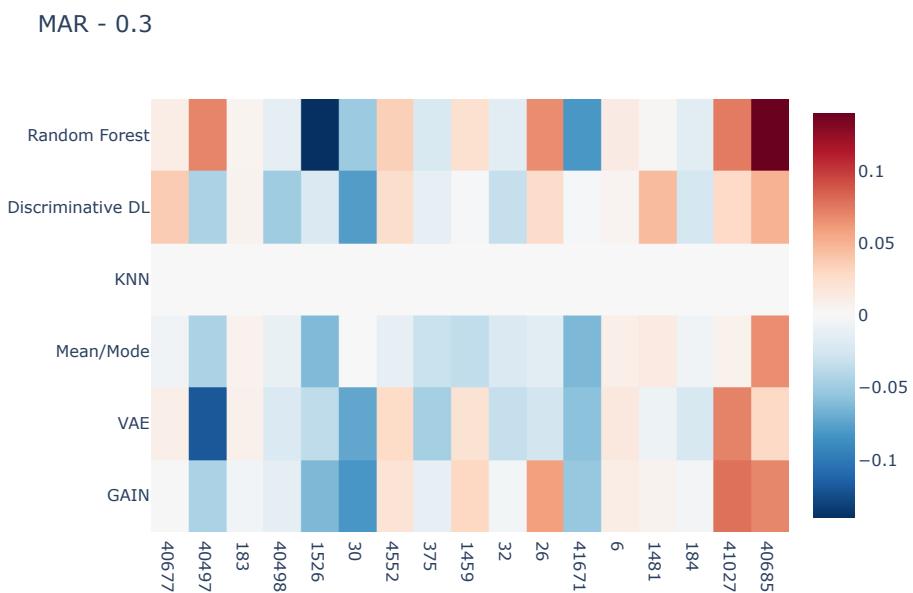


Abbildung E.11.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction

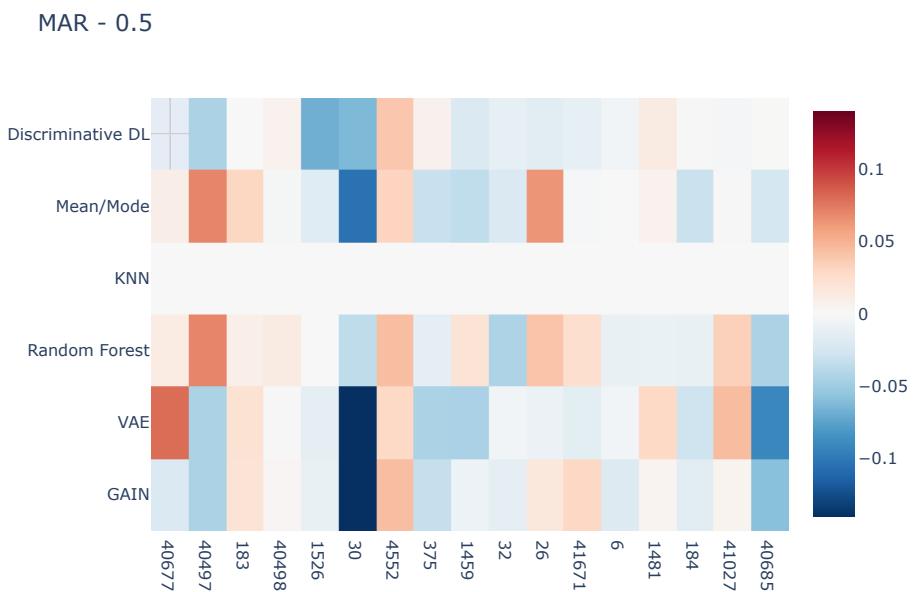


Abbildung E.12.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction

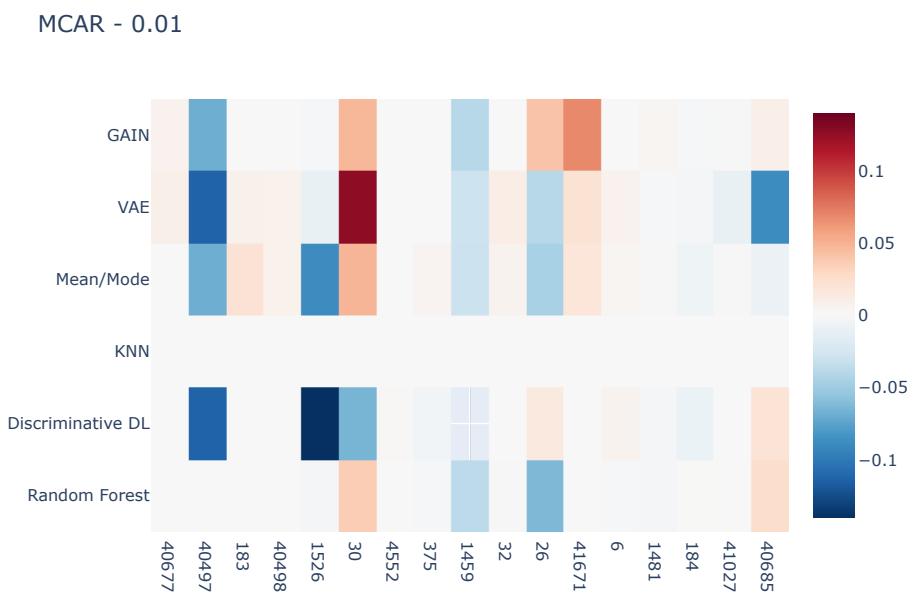


Abbildung E.13.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction

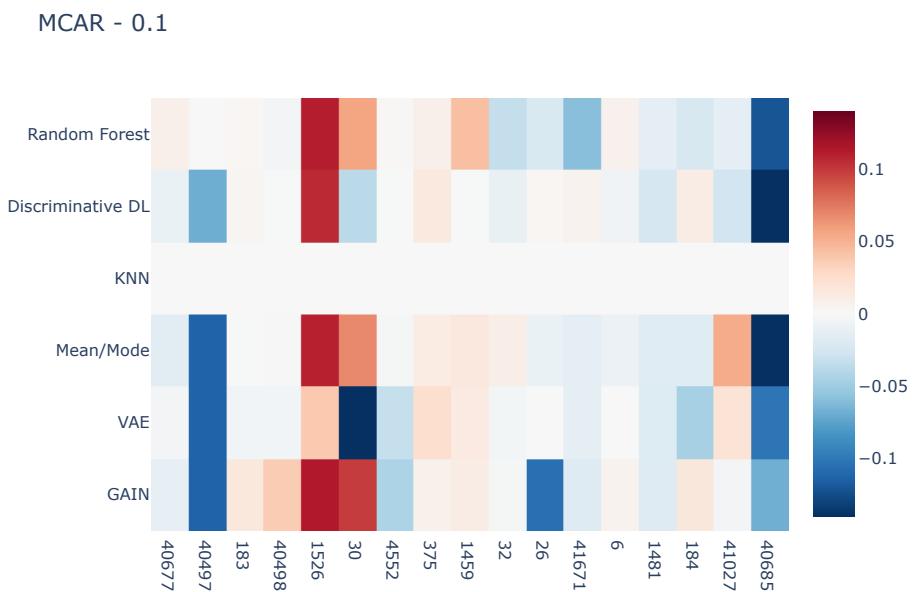


Abbildung E.14.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction

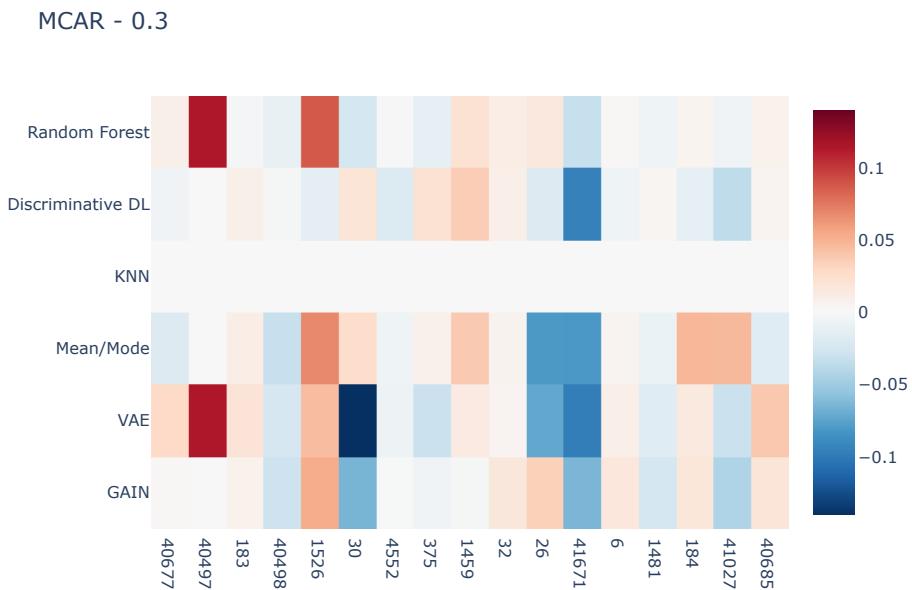


Abbildung E.15.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction

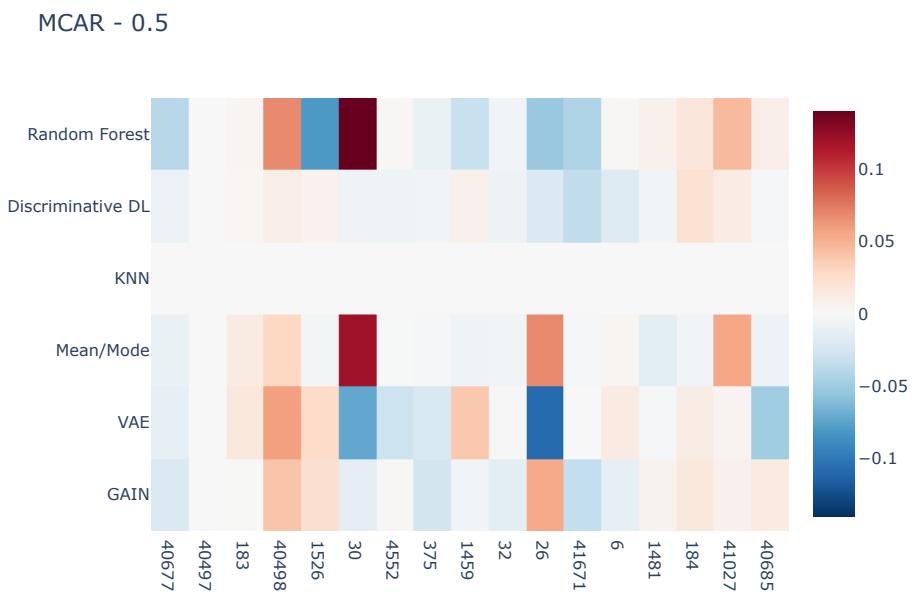


Abbildung E.16.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction

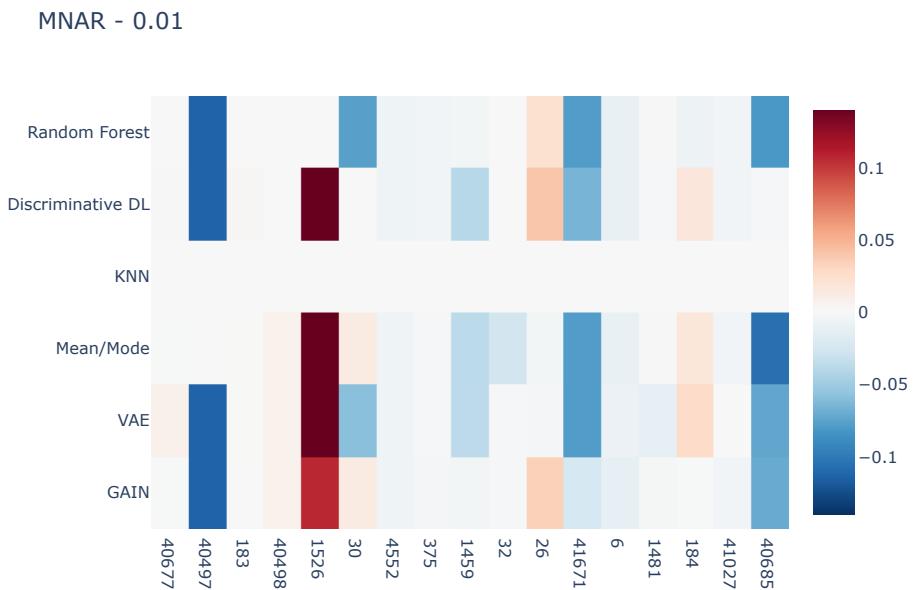


Abbildung E.17.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction

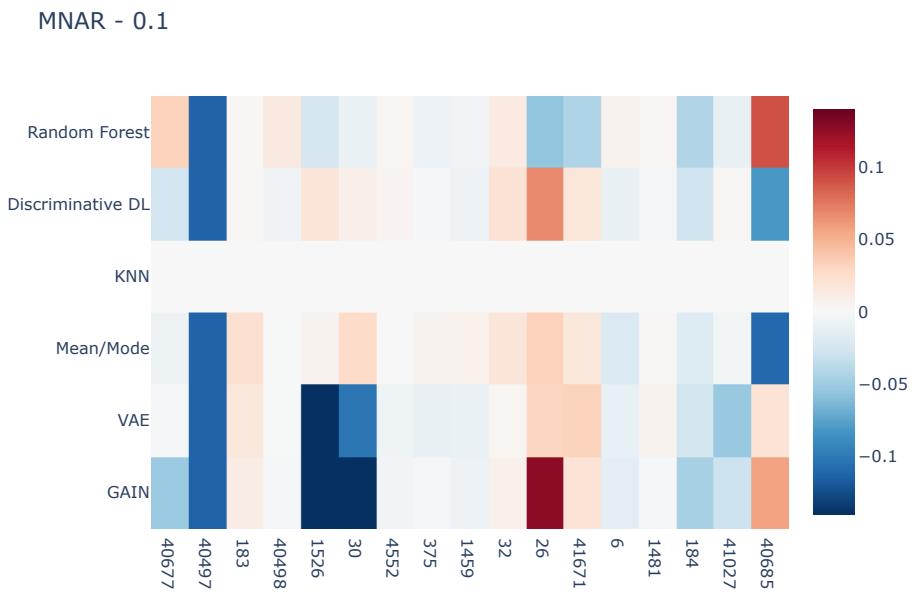


Abbildung E.18.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction

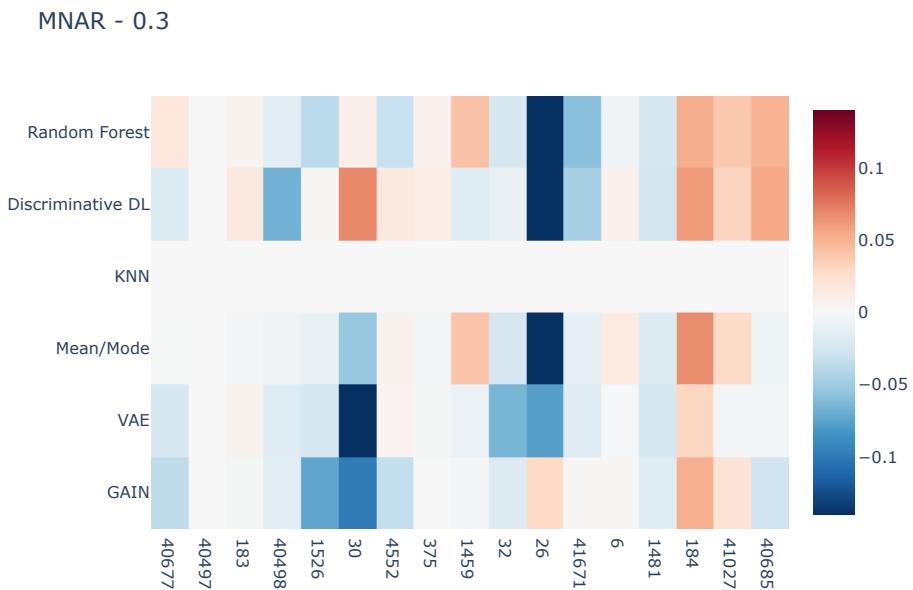


Abbildung E.19.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction

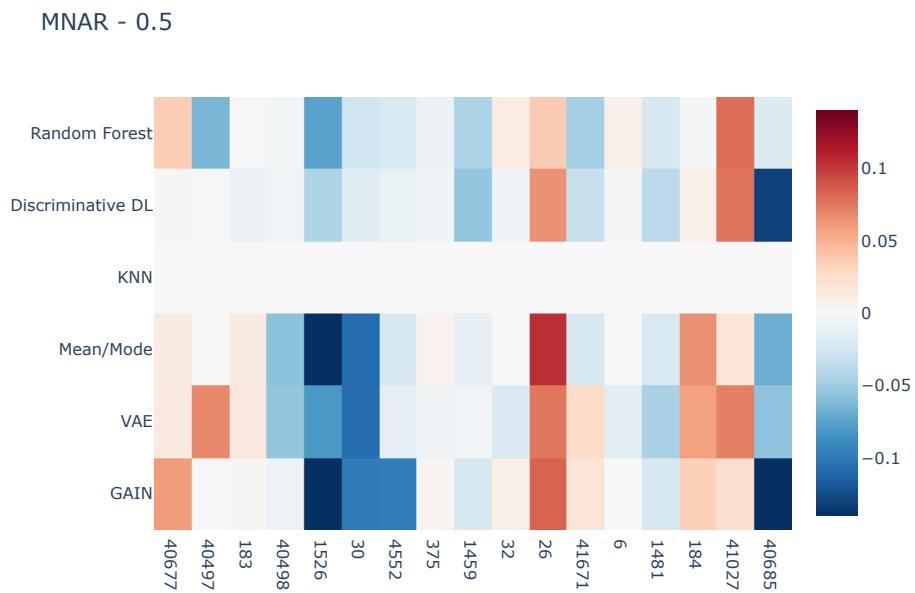


Abbildung E.20.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction

F. Appendix Regression Subset Experiments

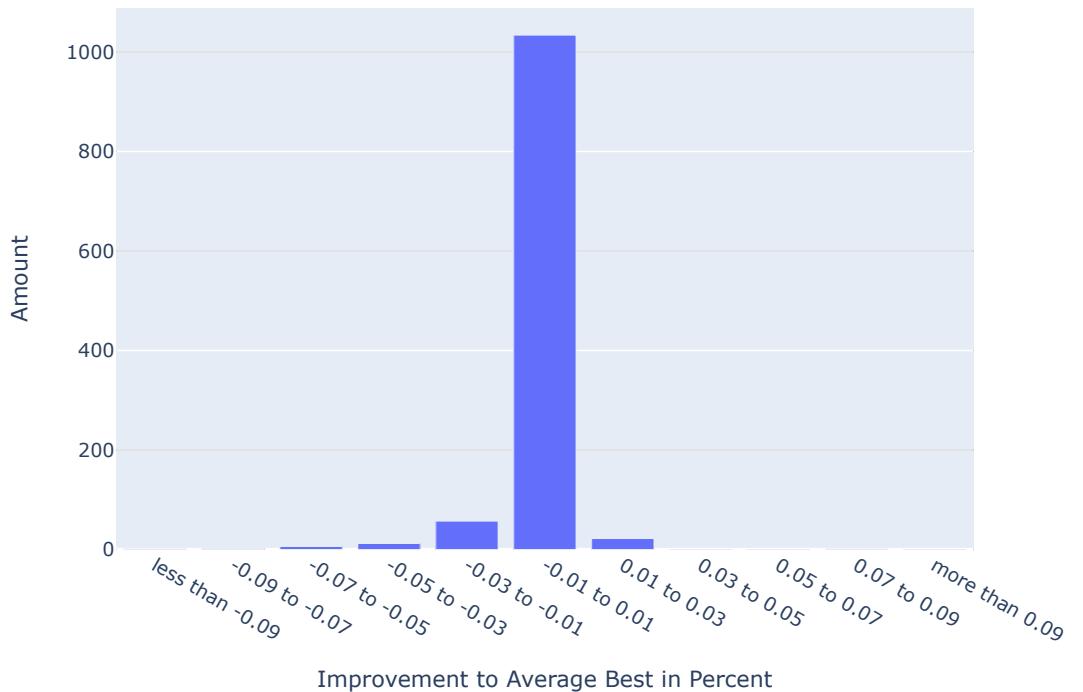


Abbildung F.1.: Improvement for Experiments Relative to the Average Best Imputation Method by Percentage

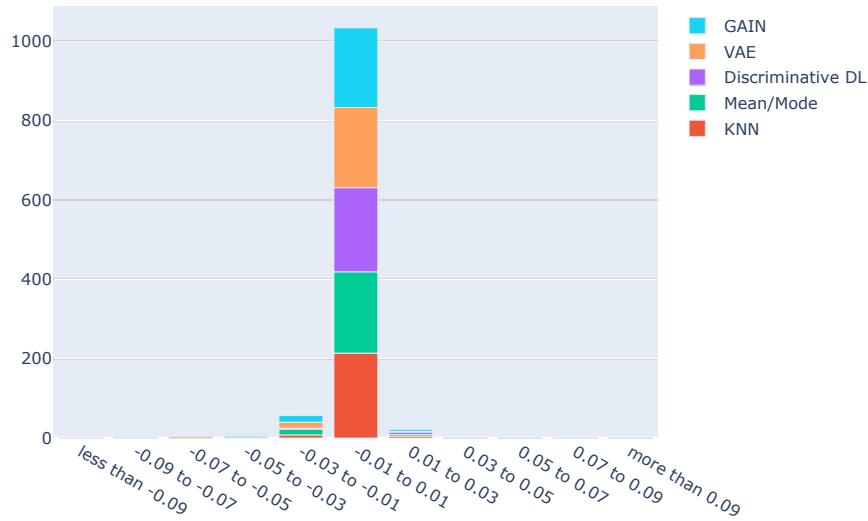


Abbildung F.2.: Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Imputation Method

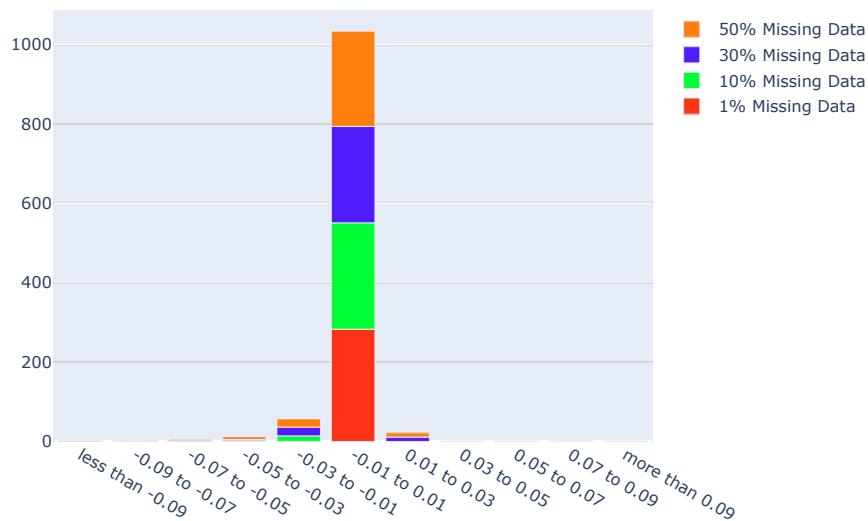


Abbildung F.3.: Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Missing Fraction

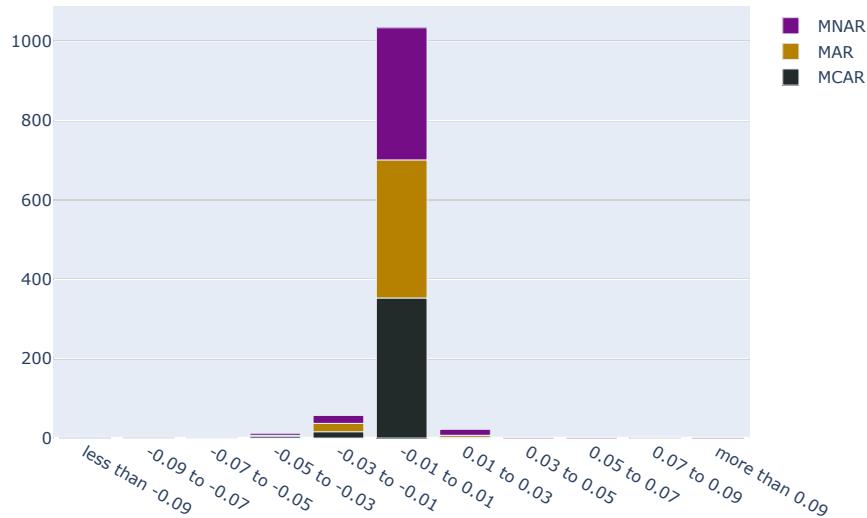


Abbildung F.4.: Improvement for Experiments Relative to the Average Best Imputation Method by Percentage by Missingness Pattern

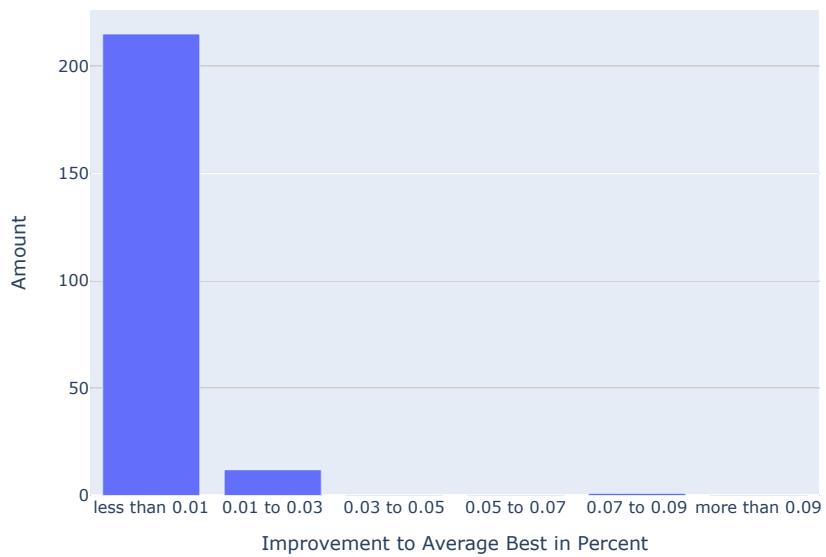


Abbildung F.5.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage

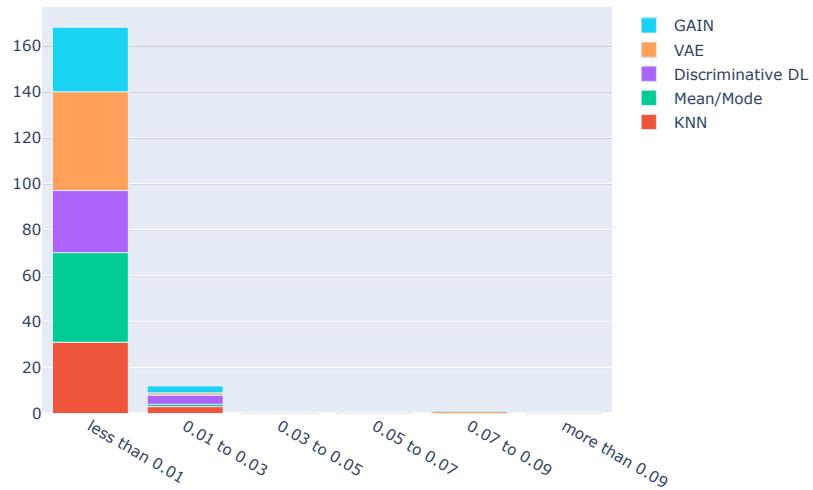


Abbildung F.6.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Imputation Method

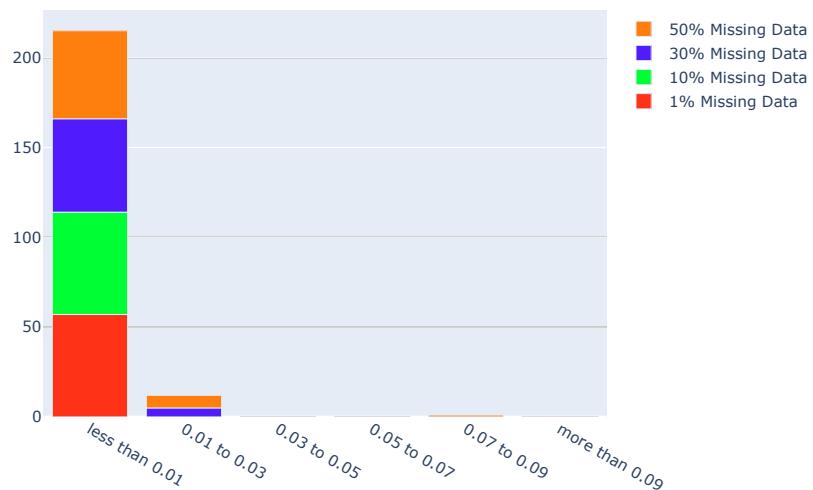


Abbildung F.7.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Missing Fraction

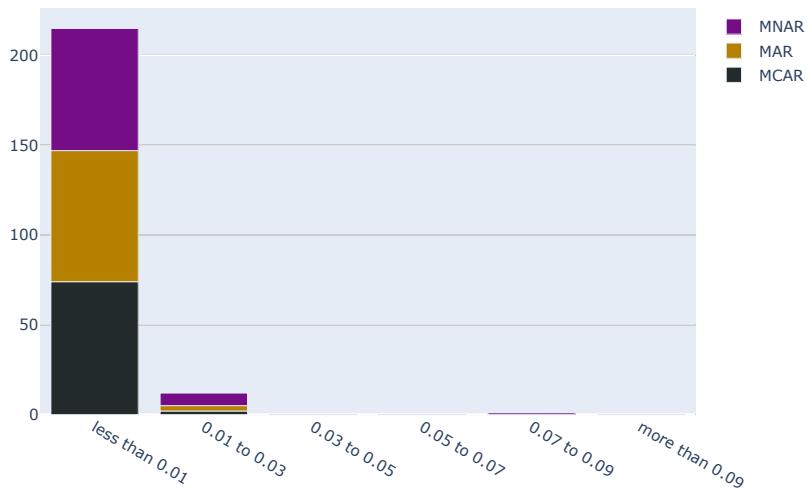


Abbildung F.8.: Improvement for Best Imputation Method per Experiment Scenario Relative to the Average Best Imputation Method by Percentage by Missingness Pattern

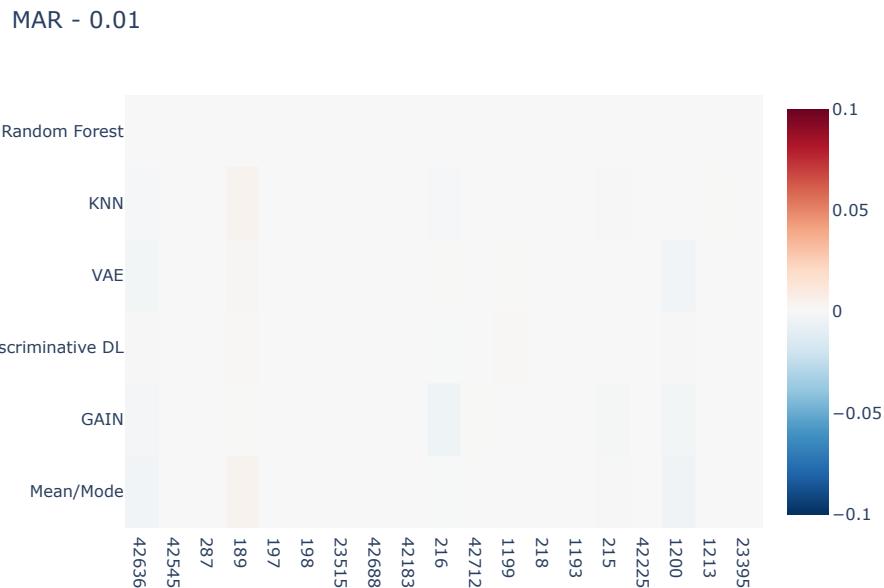


Abbildung F.9.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.01 Missing Fraction

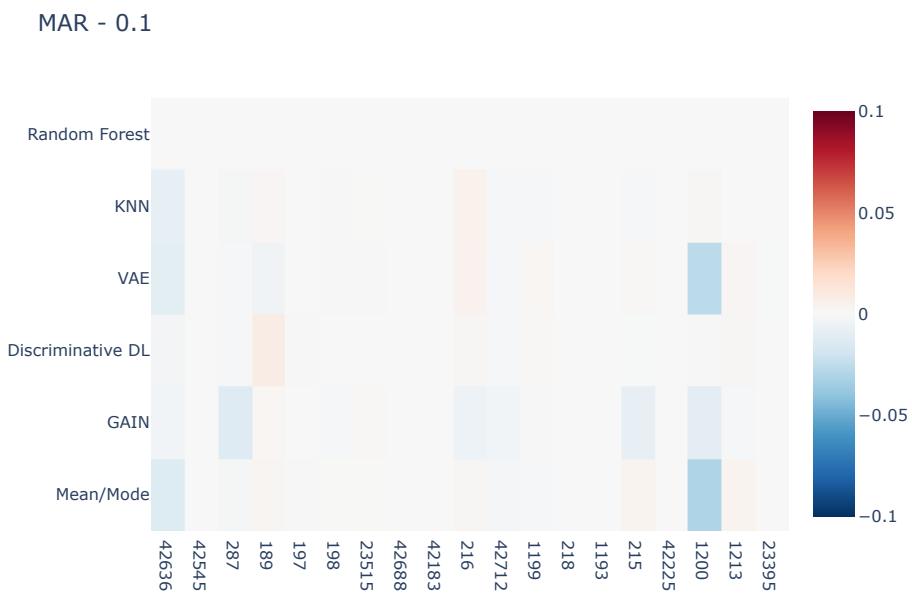


Abbildung F.10.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.1 Missing Fraction

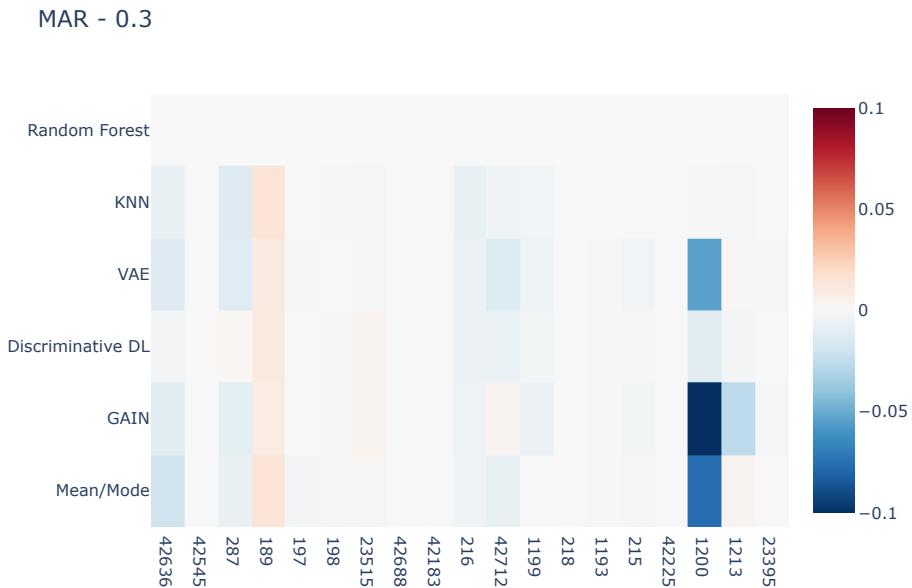


Abbildung F.11.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.3 Missing Fraction

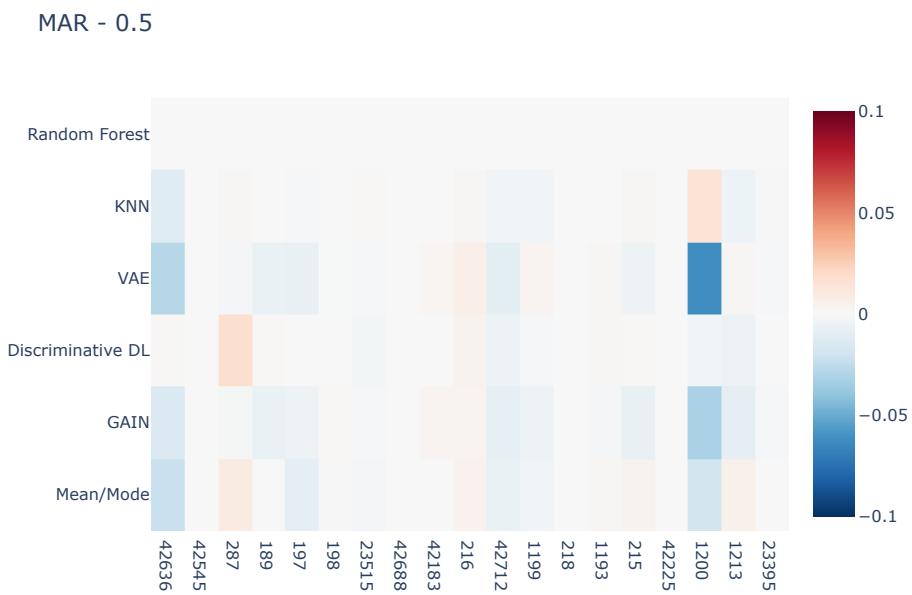


Abbildung F.12.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MAR and 0.5 Missing Fraction

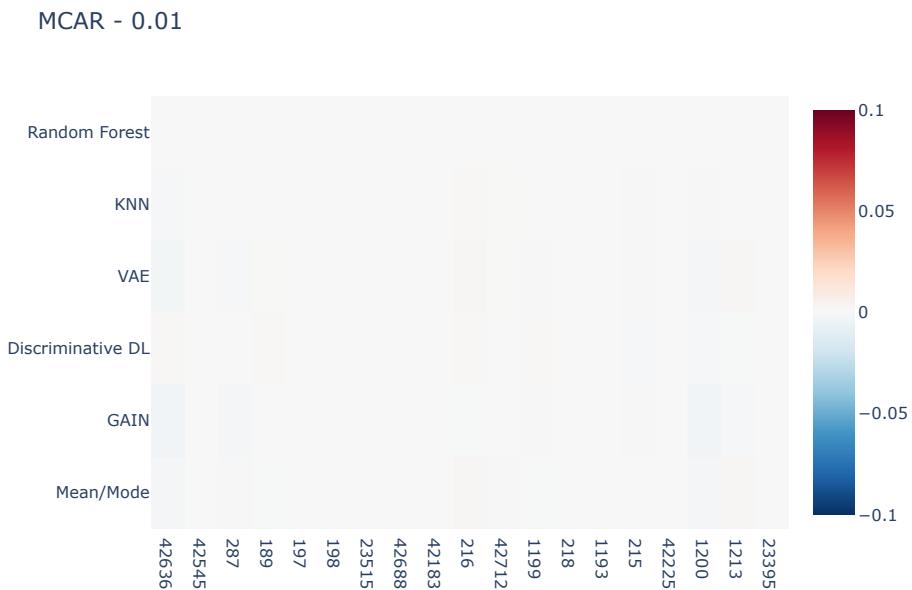


Abbildung F.13.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.01 Missing Fraction

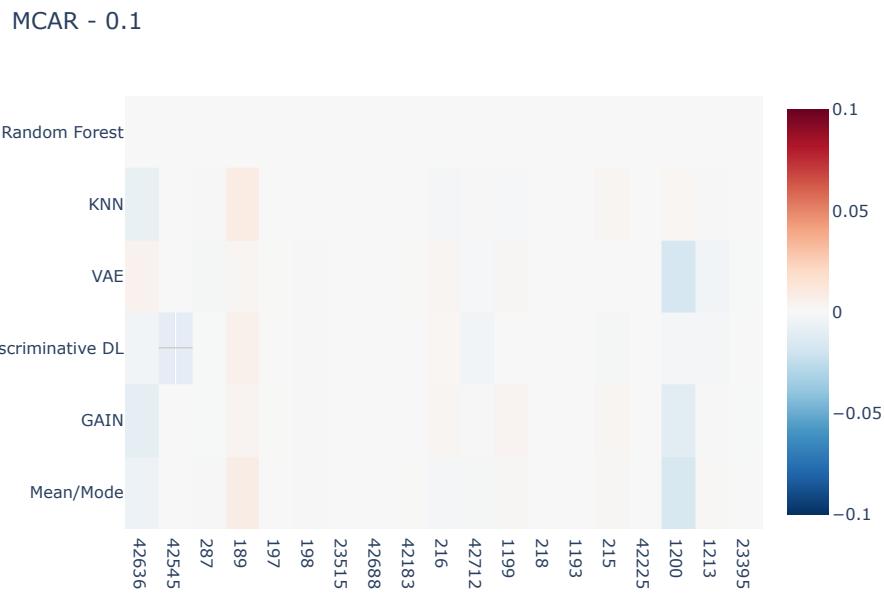


Abbildung F.14.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.1 Missing Fraction

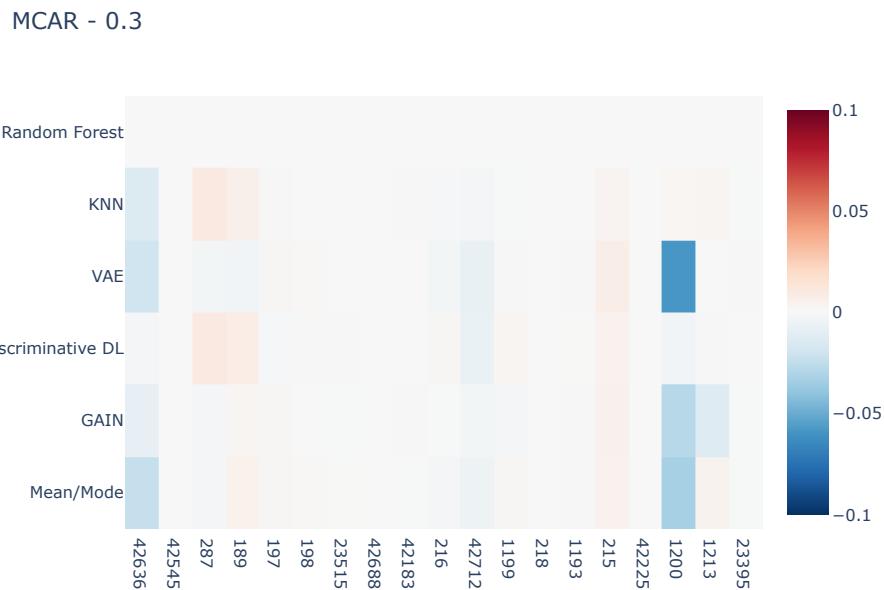


Abbildung F.15.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.3 Missing Fraction

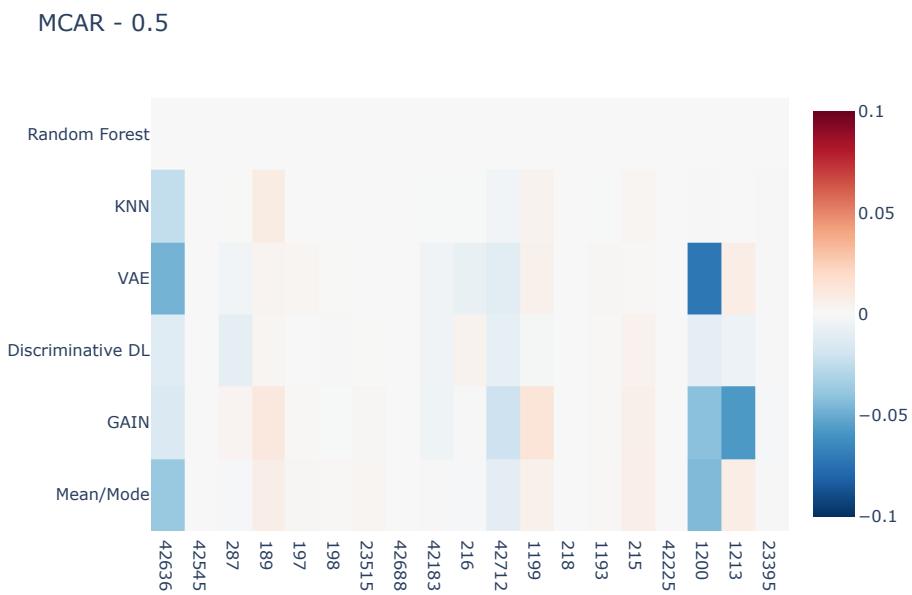


Abbildung F.16.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MCAR and 0.5 Missing Fraction

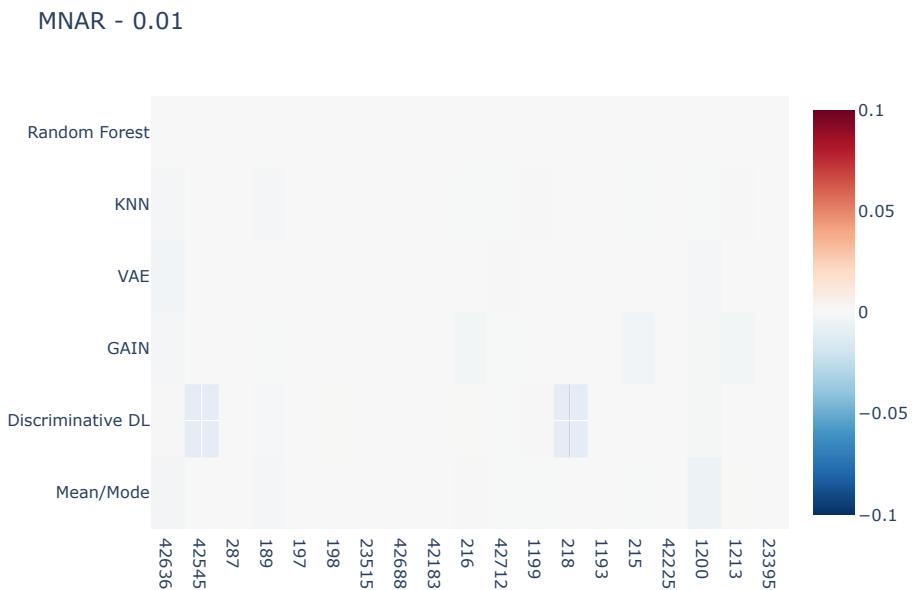


Abbildung F.17.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.01 Missing Fraction

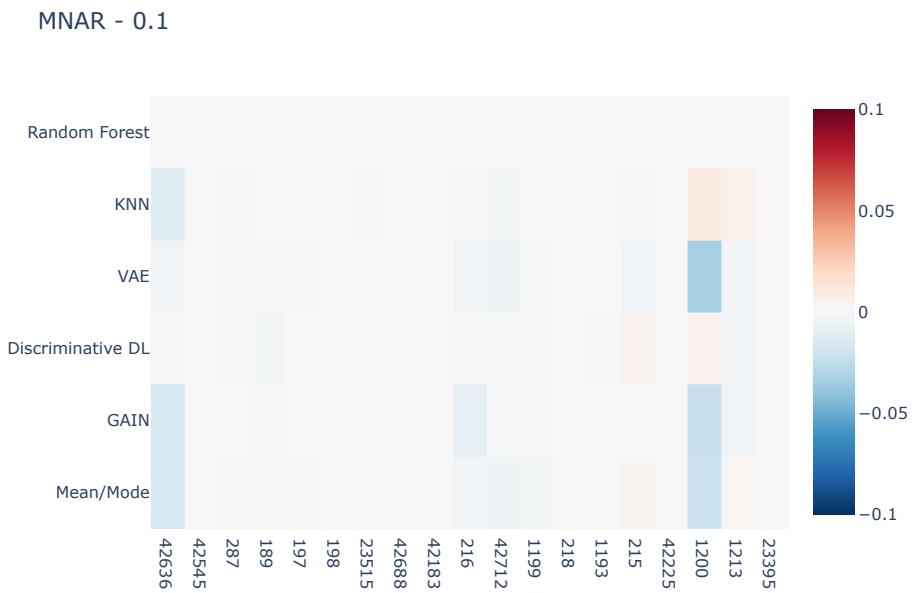


Abbildung F.18.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.1 Missing Fraction

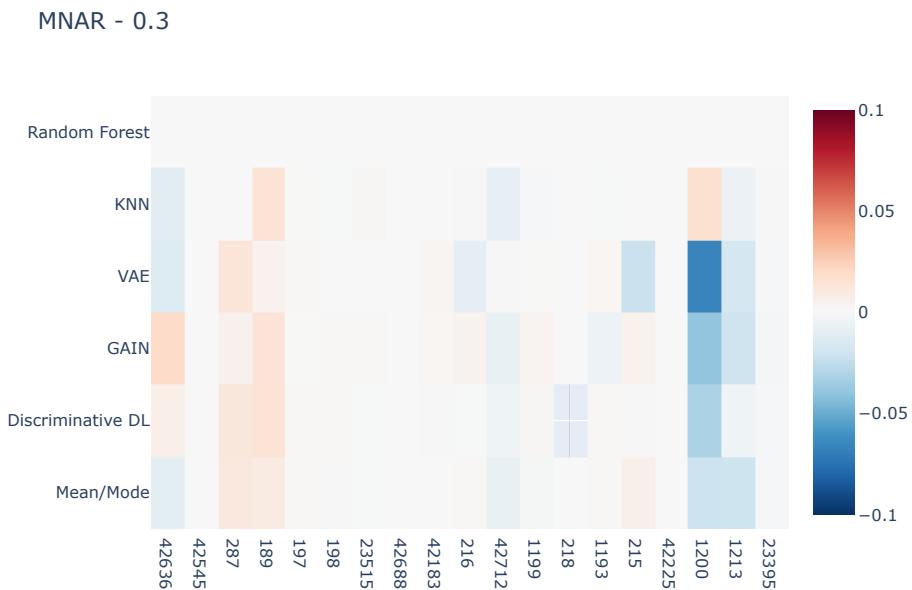


Abbildung F.19.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.3 Missing Fraction

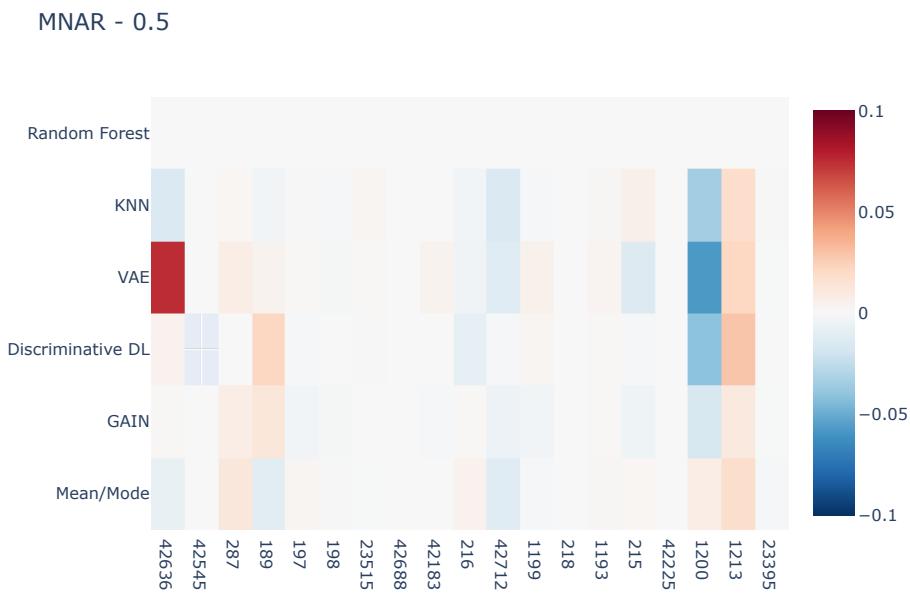


Abbildung F.20.: Improvements for all Imputation Methods Relative to Average Best Method per Dataset for MNAR and 0.5 Missing Fraction

G. Appendix Binary Classification Baseline - Imputed Comparisons

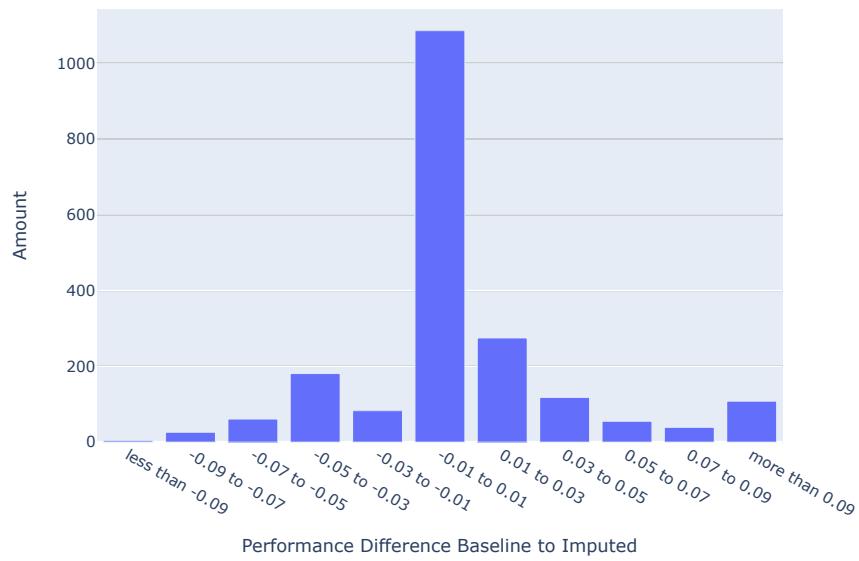


Abbildung G.1.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data

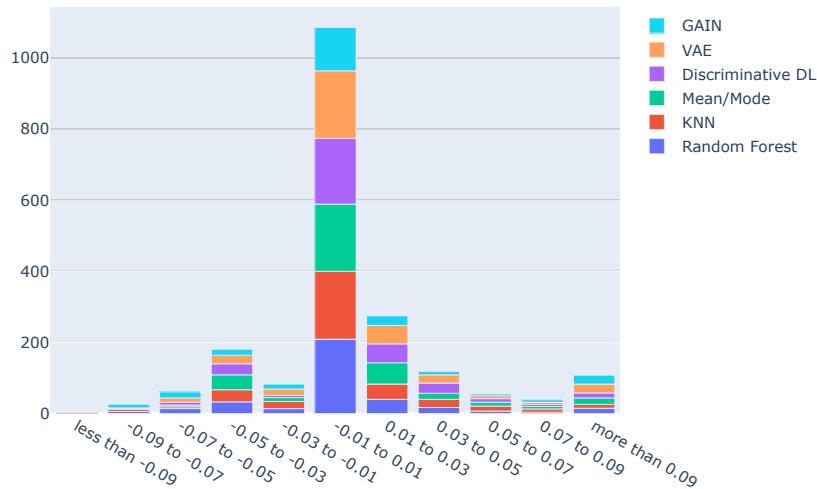


Abbildung G.2.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Imputation Method

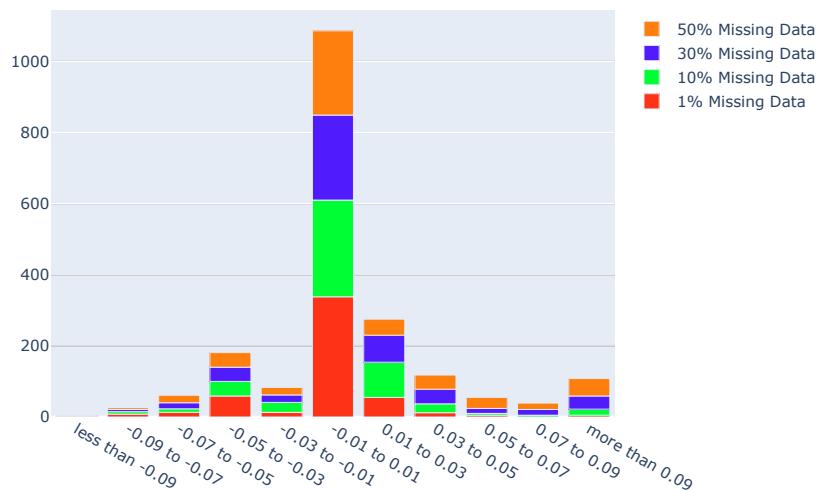


Abbildung G.3.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Fraction

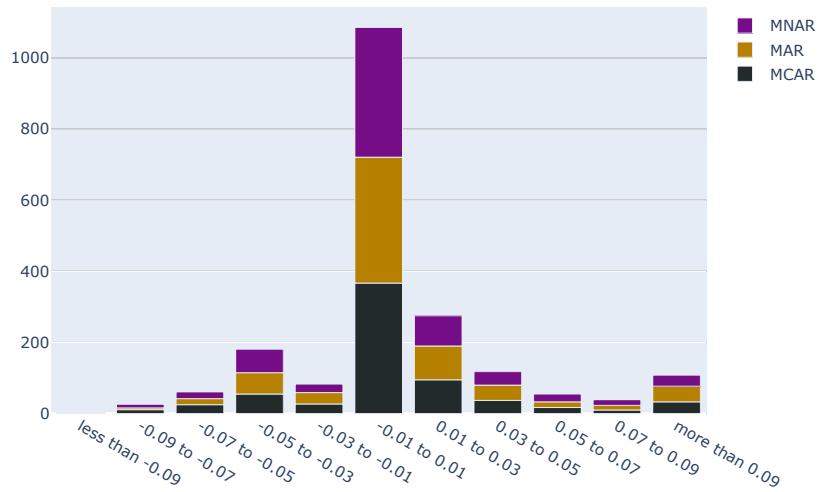


Abbildung G.4.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Pattern

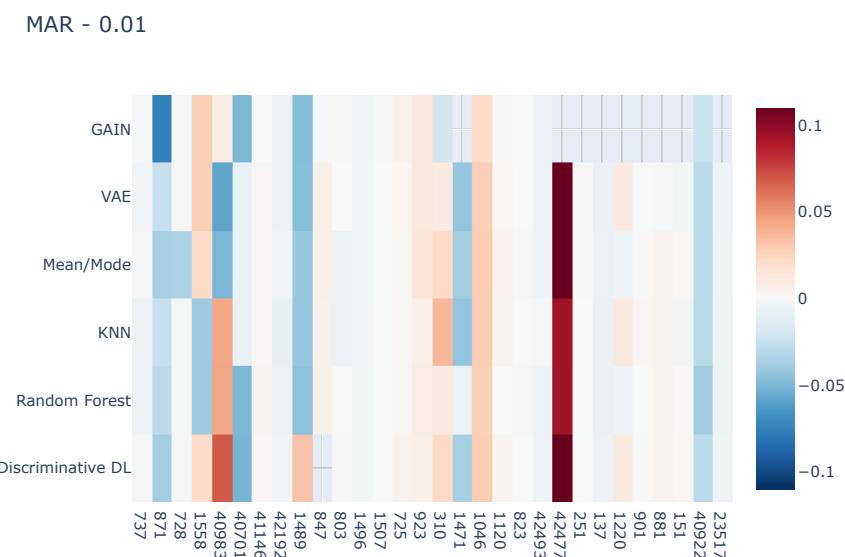


Abbildung G.5.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.01 Missing Fraction

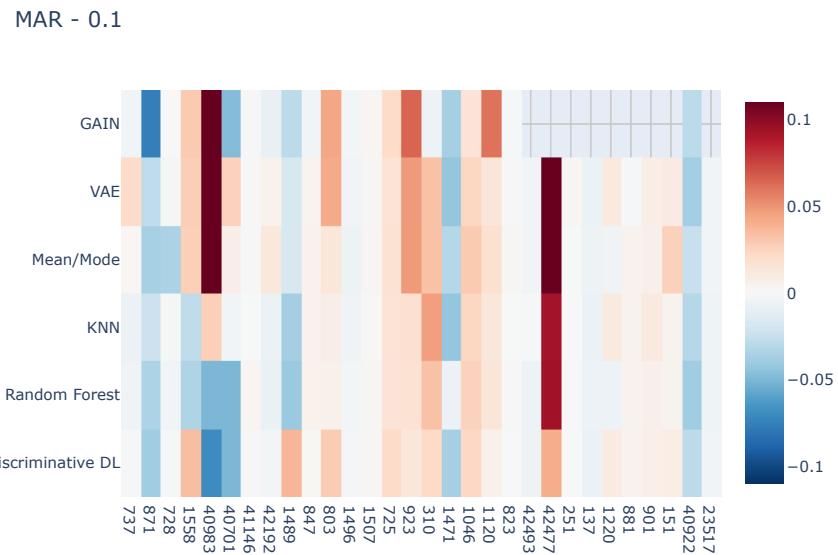


Abbildung G.6.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.1 Missing Fraction

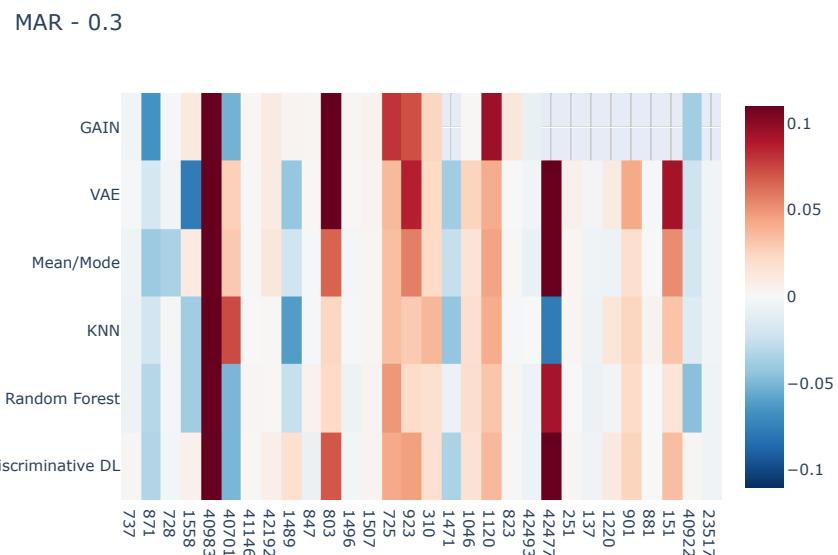


Abbildung G.7.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.3 Missing Fraction

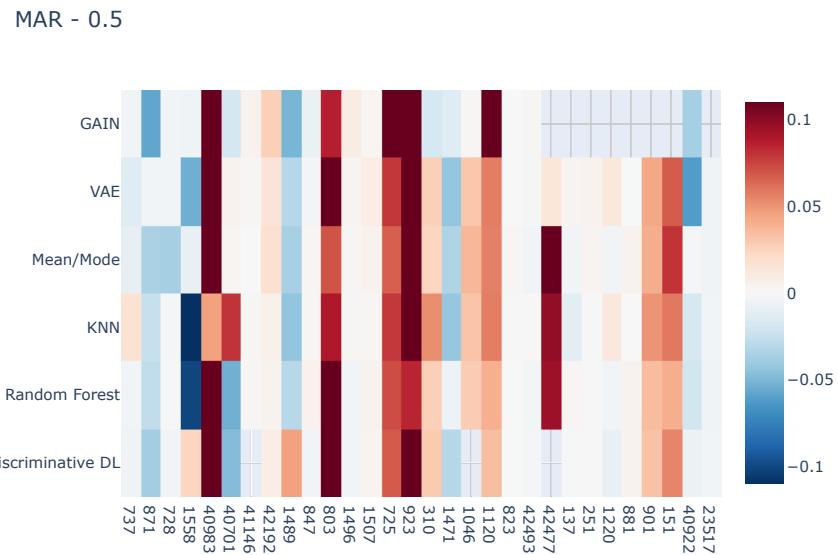


Abbildung G.8.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.5 Missing Fraction

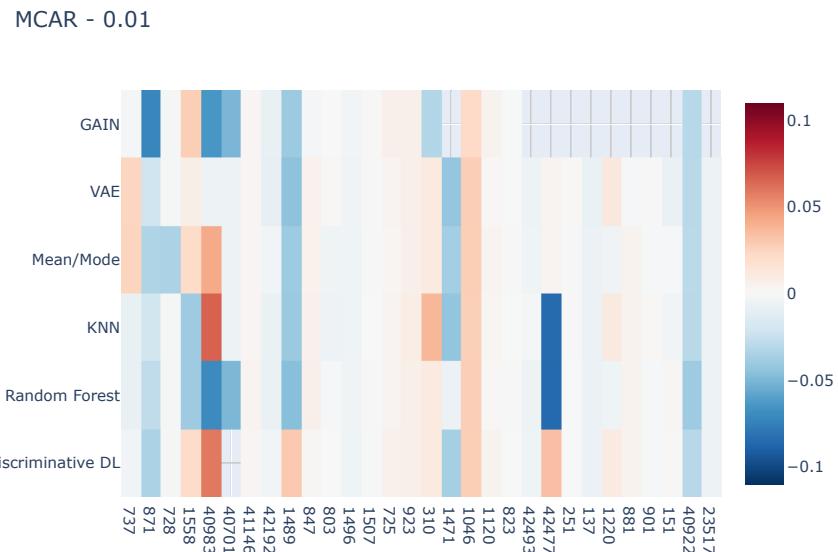


Abbildung G.9.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.01 Missing Fraction

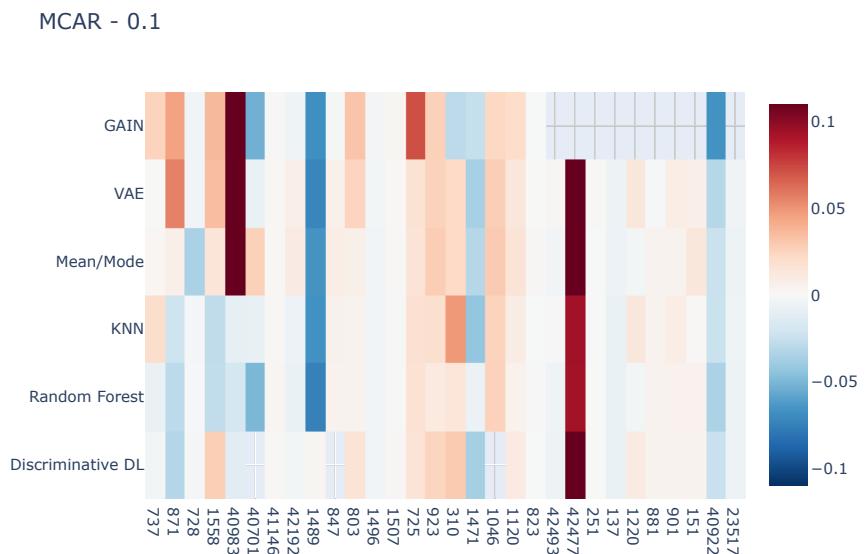


Abbildung G.10.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.1 Missing Fraction

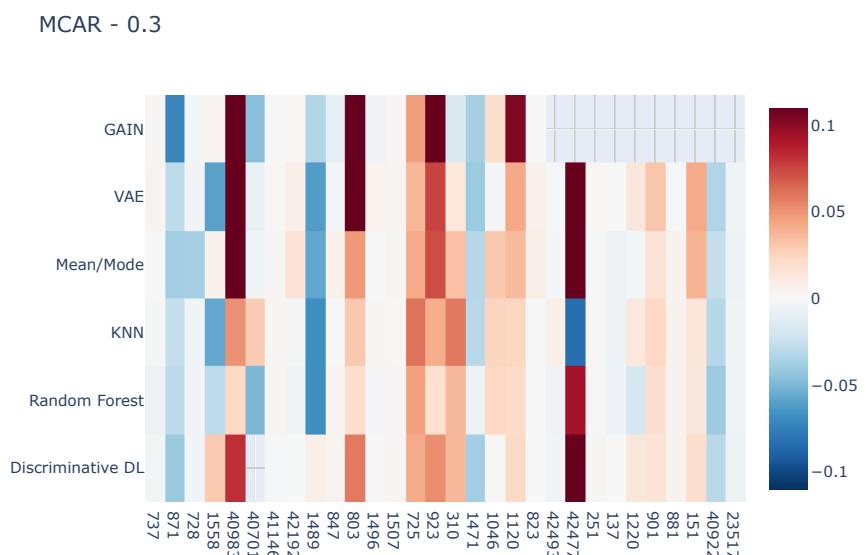


Abbildung G.11.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.3 Missing Fraction

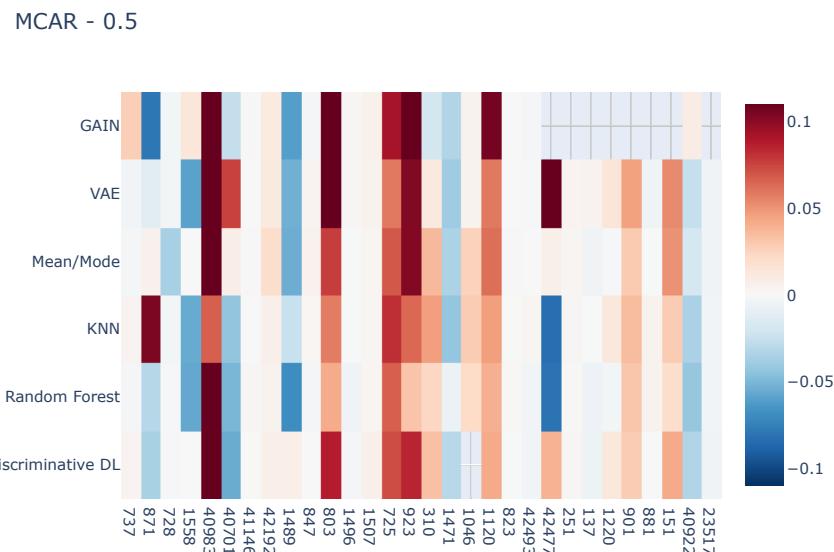


Abbildung G.12.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.5 Missing Fraction

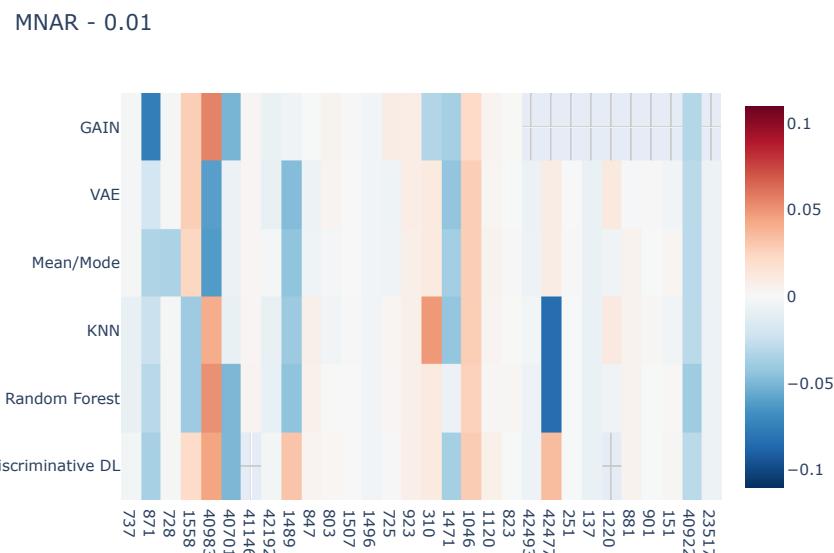


Abbildung G.13.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.01 Missing Fraction

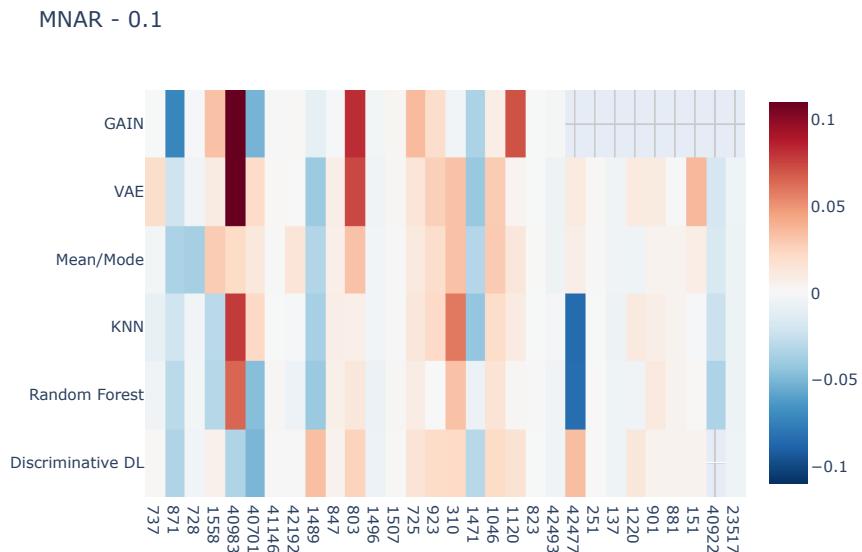


Abbildung G.14.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.1 Missing Fraction

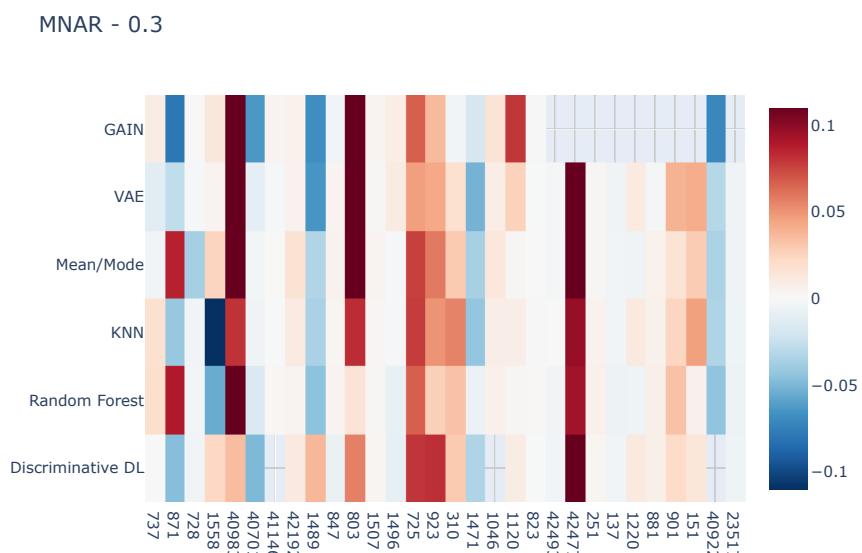


Abbildung G.15.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.3 Missing Fraction

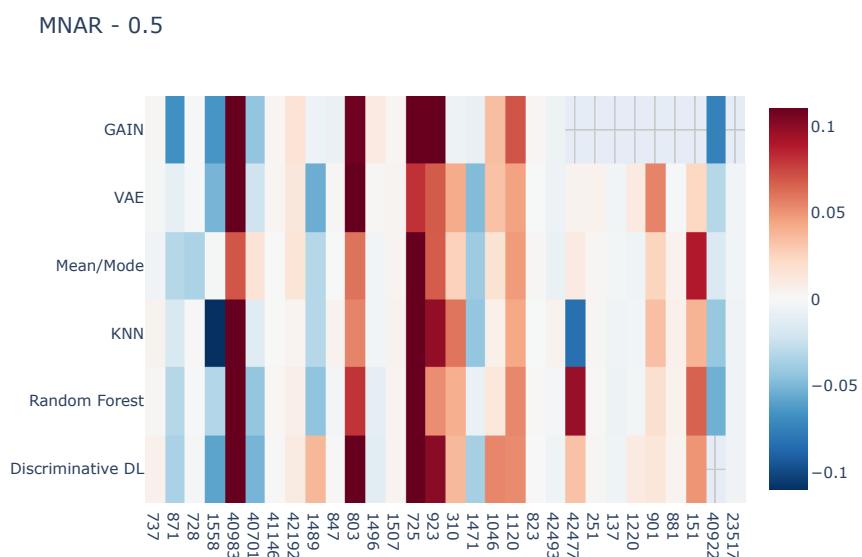


Abbildung G.16.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.5 Missing Fraction

H. Appendix Binary Classification Imputed - Subset Comparisons

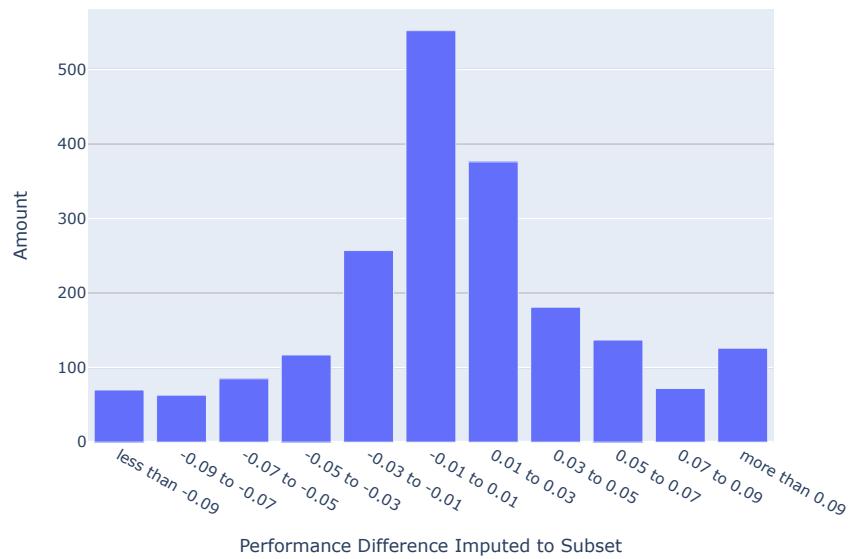


Abbildung H.1.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data

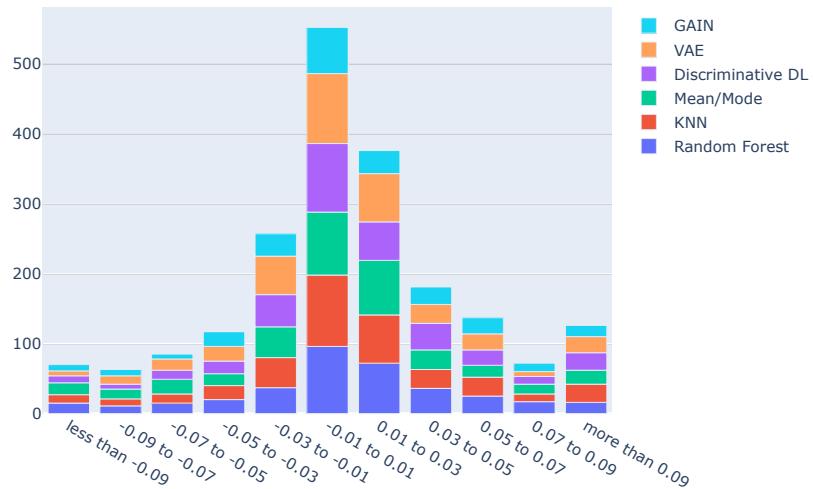


Abbildung H.2.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Imputation Method

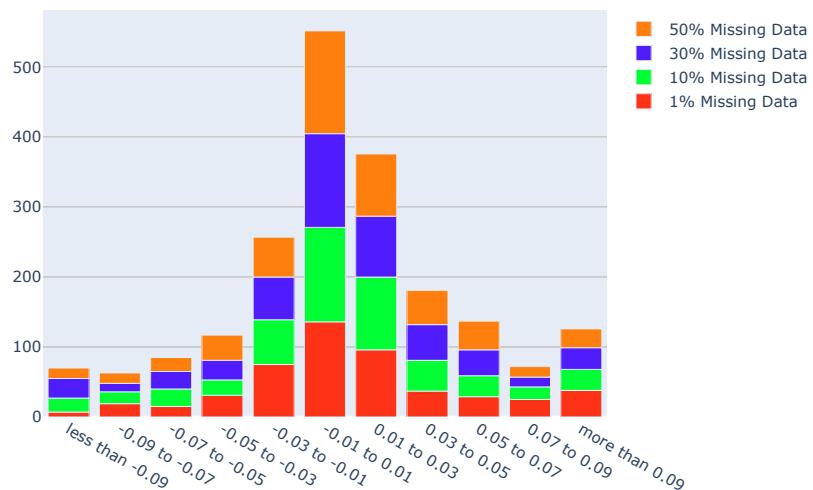


Abbildung H.3.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Fraction

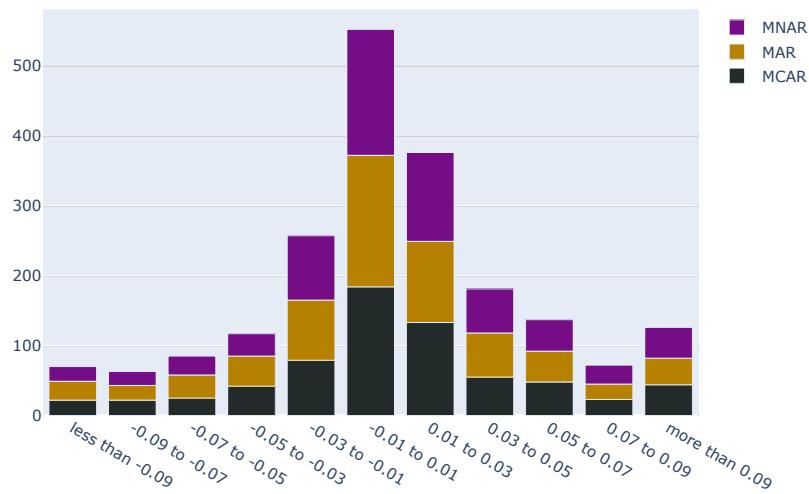


Abbildung H.4.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Pattern

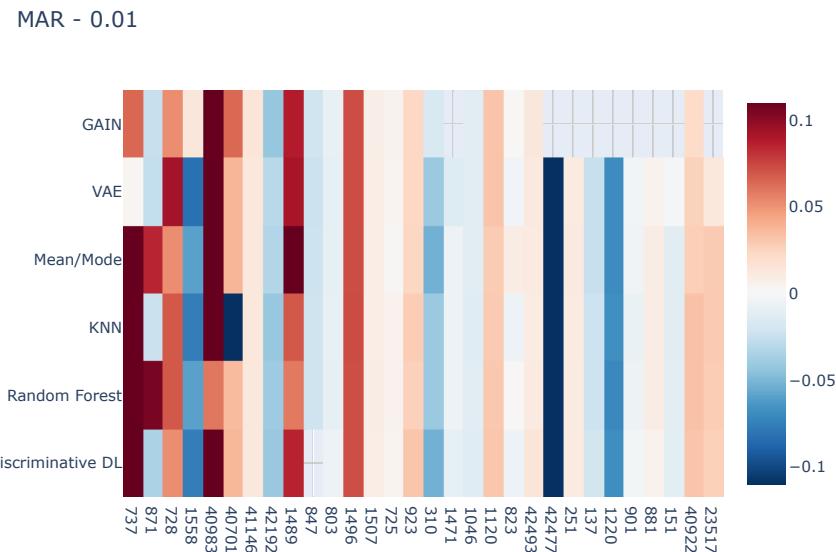


Abbildung H.5.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.01 Missing Fraction

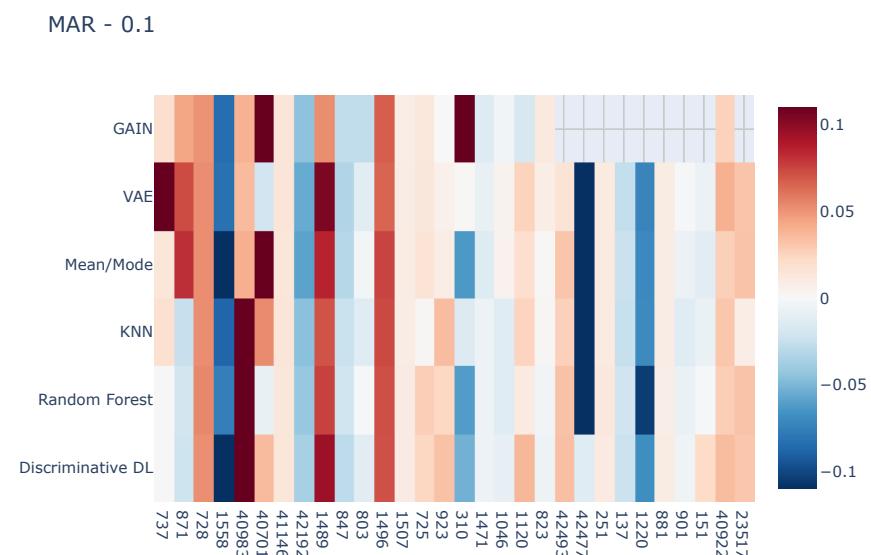


Abbildung H.6.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.1 Missing Fraction

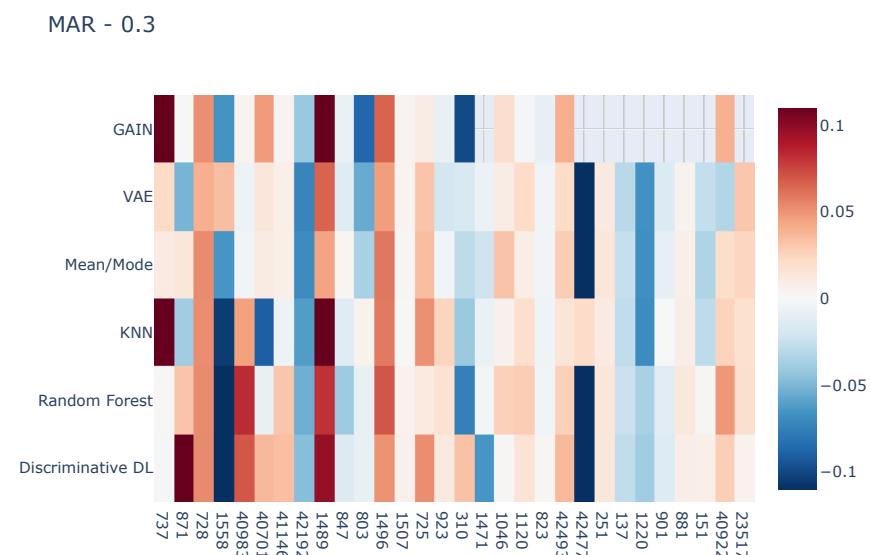


Abbildung H.7.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.3 Missing Fraction

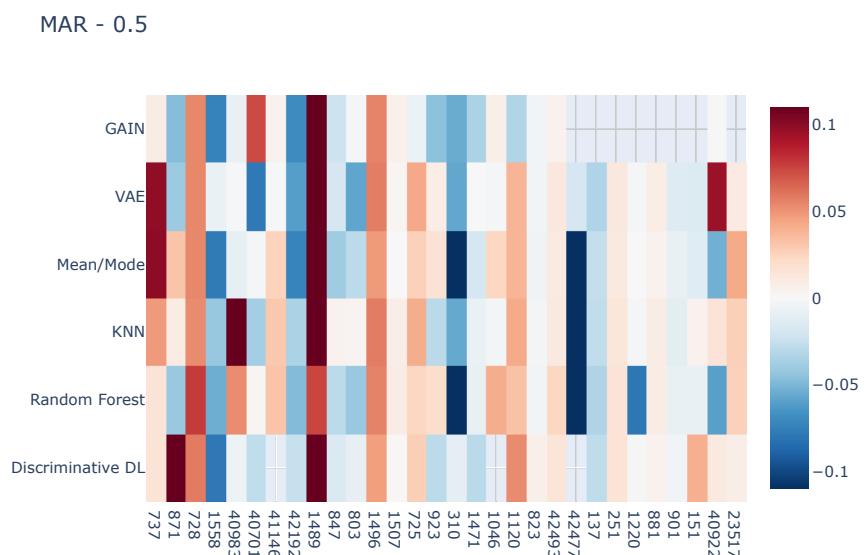


Abbildung H.8.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.5 Missing Fraction

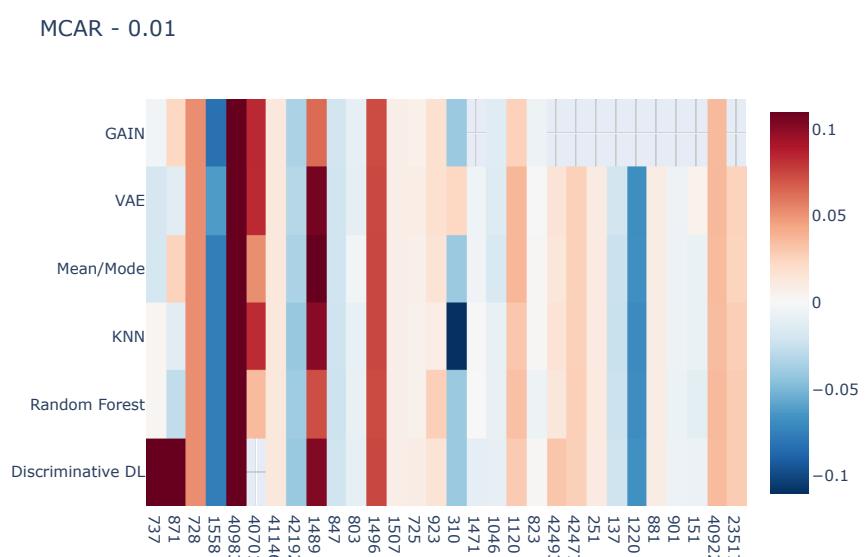


Abbildung H.9.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.01 Missing Fraction

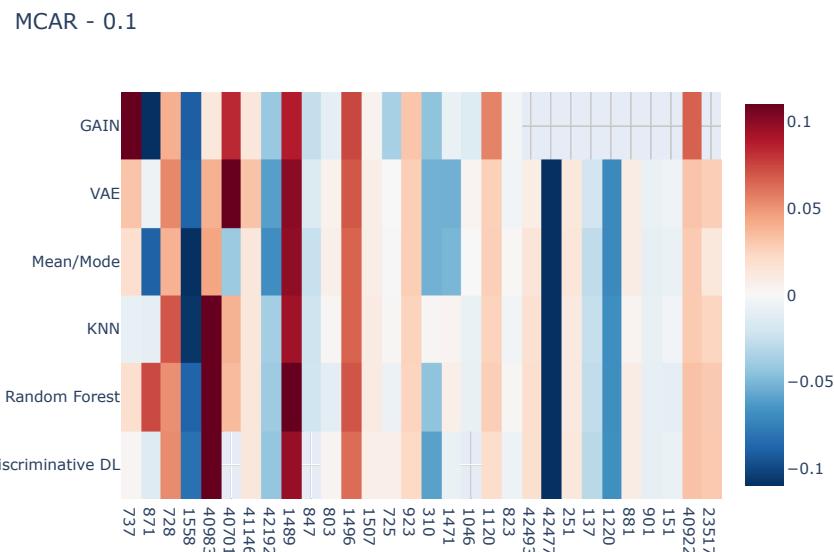


Abbildung H.10.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.1 Missing Fraction

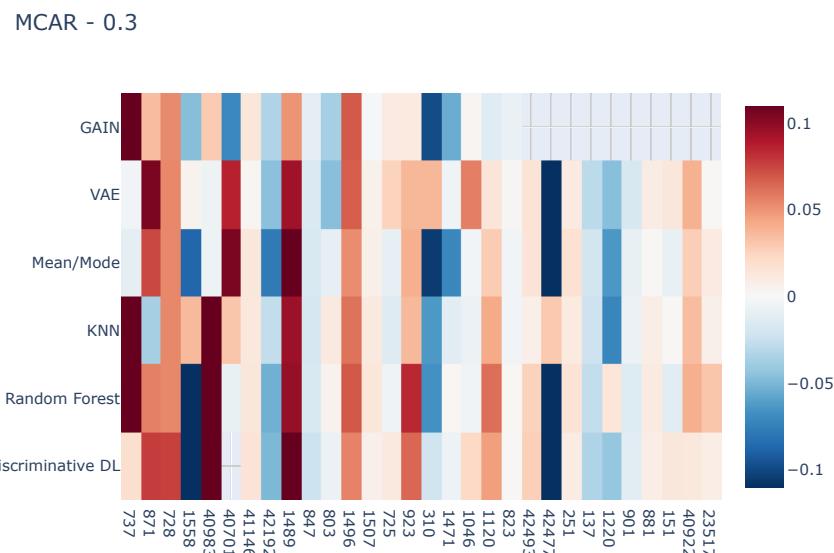


Abbildung H.11.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.3 Missing Fraction

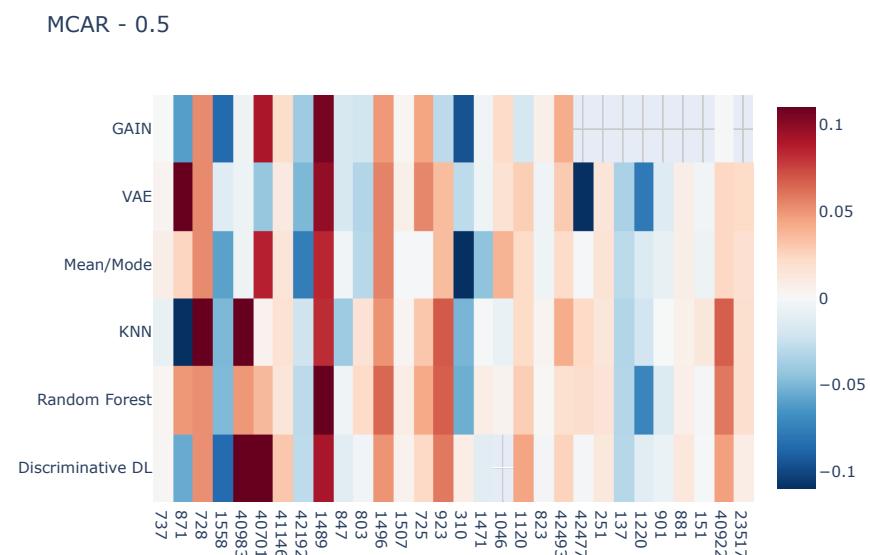


Abbildung H.12.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.5 Missing Fraction

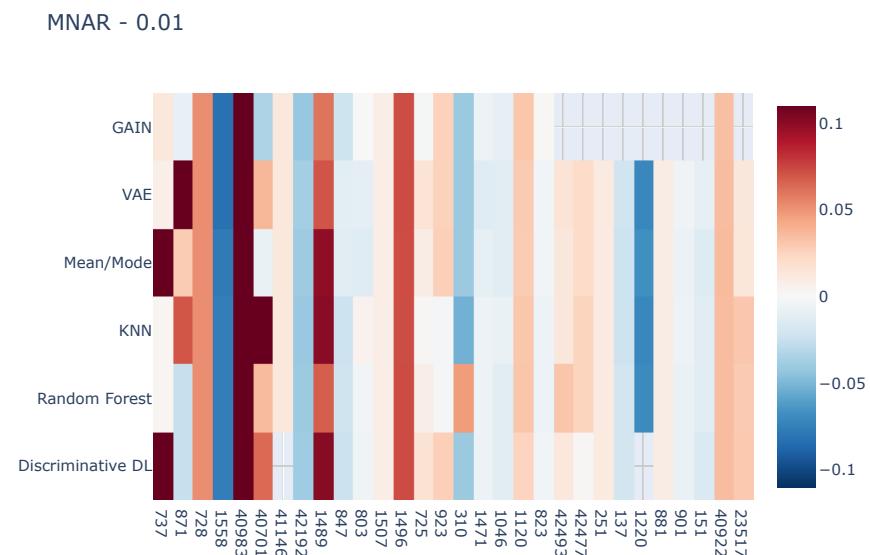


Abbildung H.13.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.01 Missing Fraction

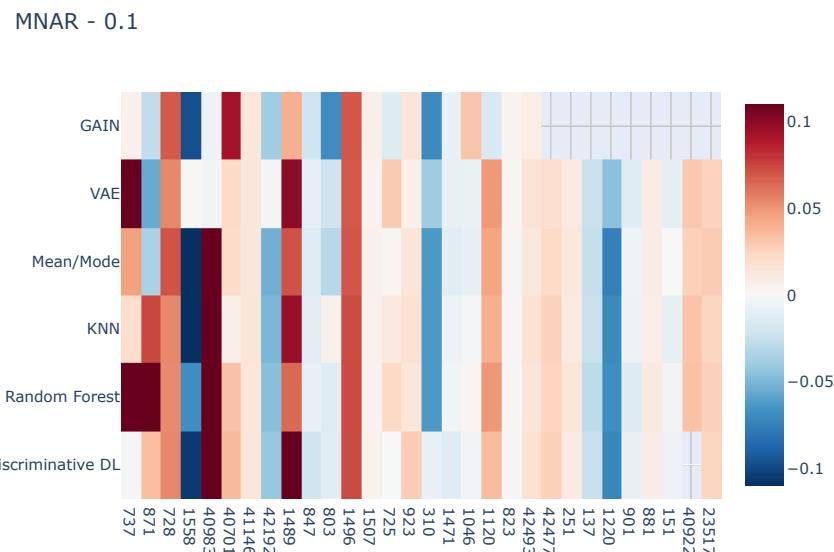


Abbildung H.14.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.1 Missing Fraction

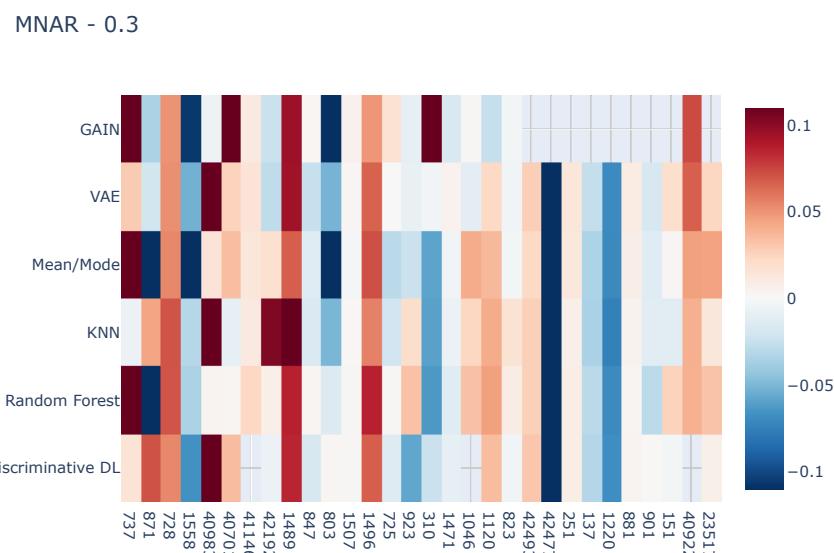


Abbildung H.15.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.3 Missing Fraction

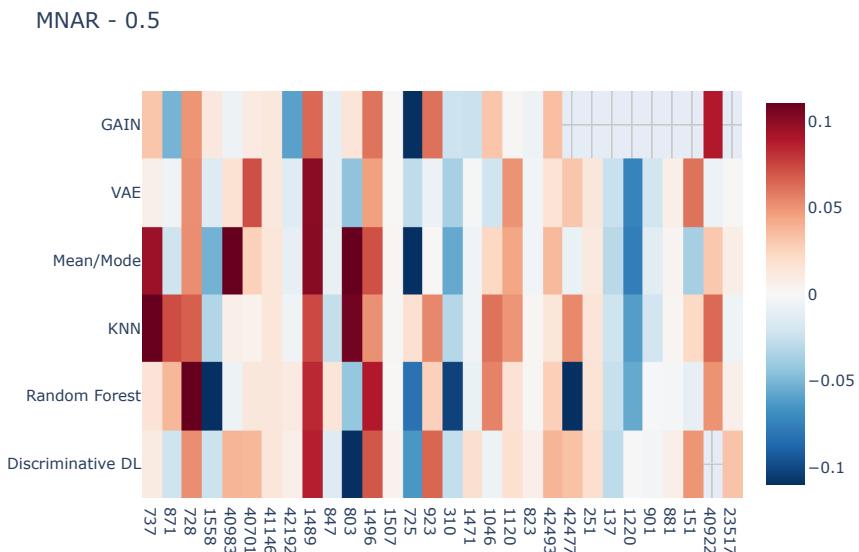


Abbildung H.16.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.5 Missing Fraction

I. Appendix Multiclass Classification Baseline - Imputed Comparisons

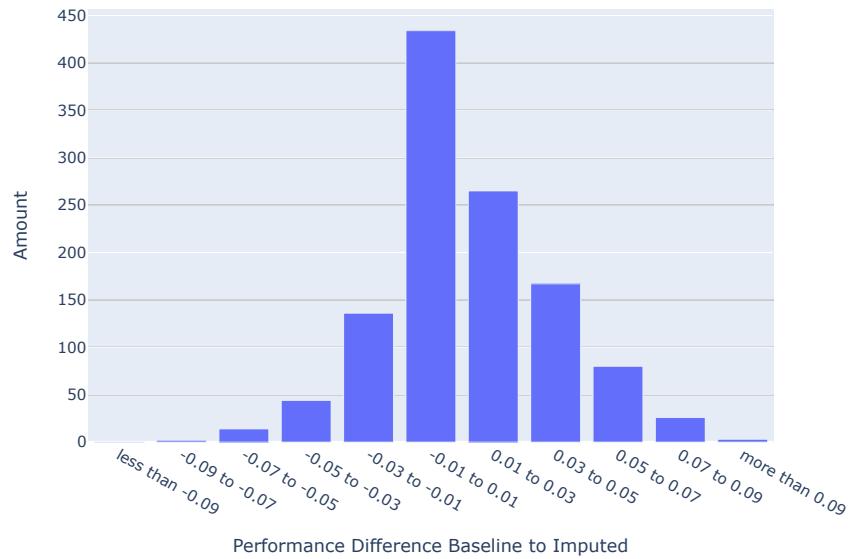


Abbildung I.1.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data

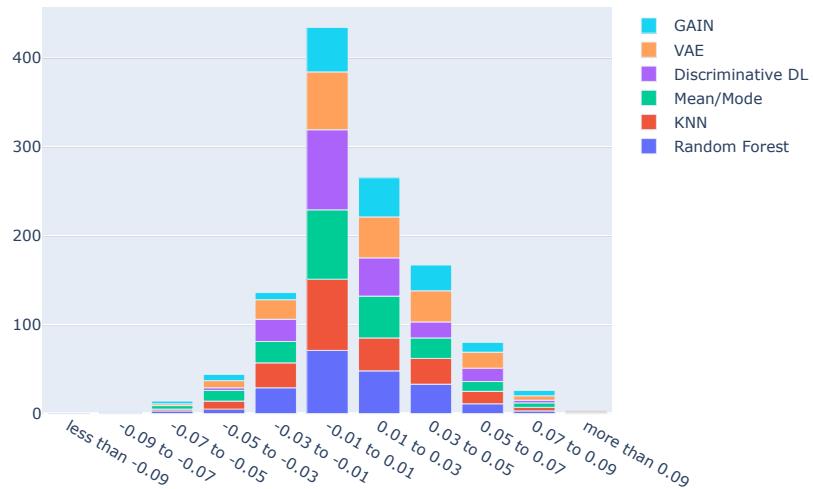


Abbildung I.2.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Imputation Method

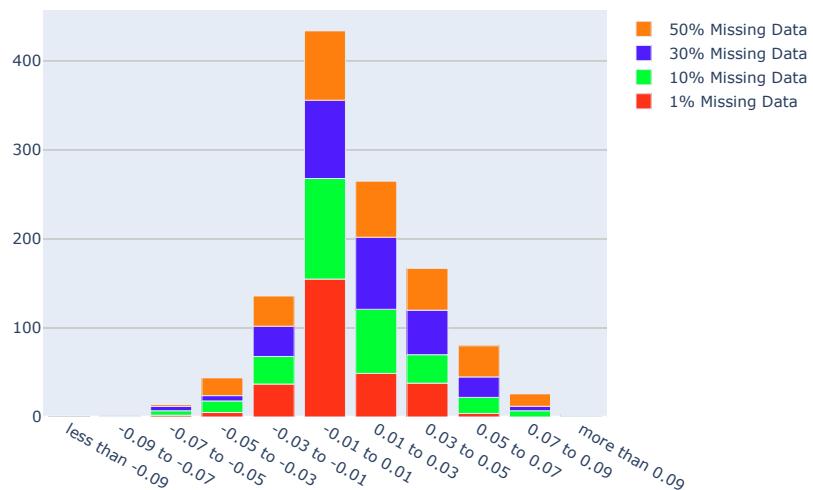


Abbildung I.3.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Fraction

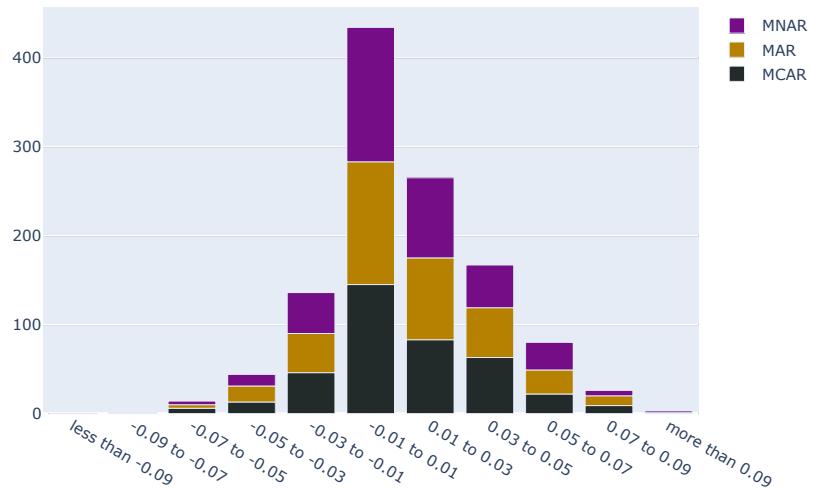


Abbildung I.4.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Pattern

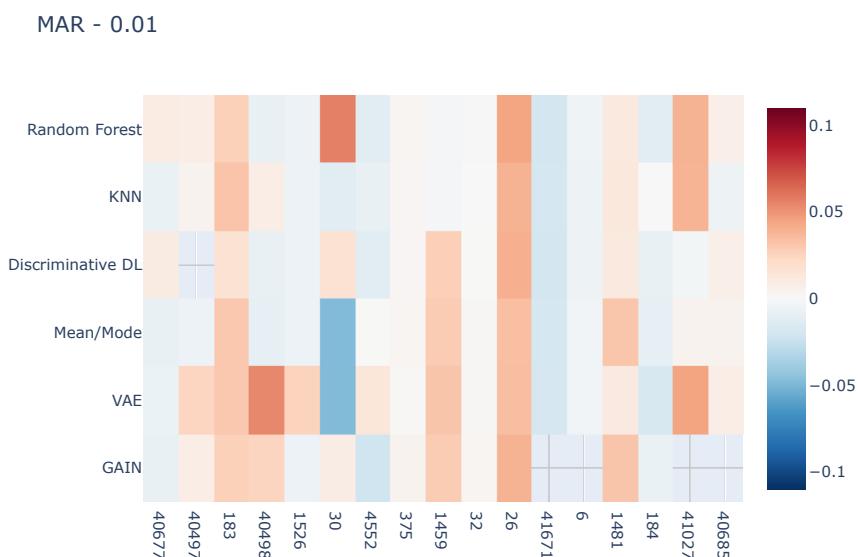


Abbildung I.5.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.01 Missing Fraction

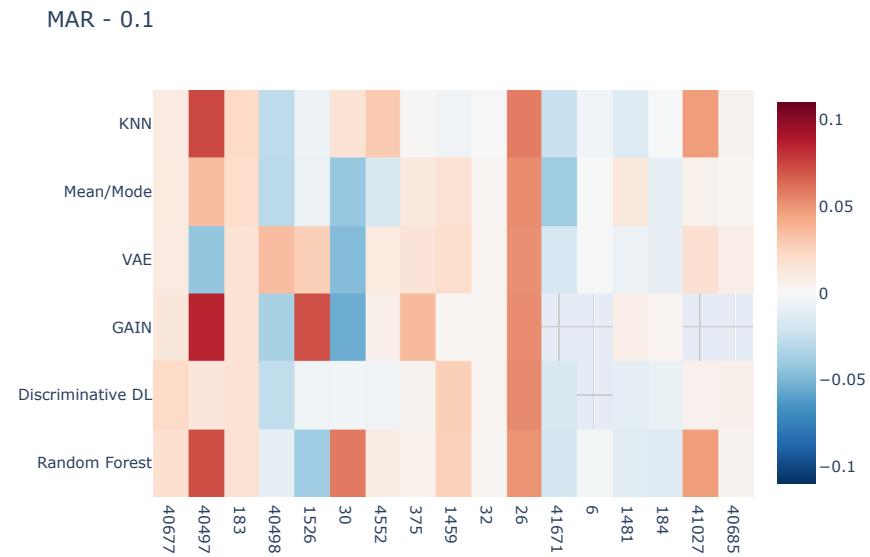


Abbildung I.6.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.1 Missing Fraction

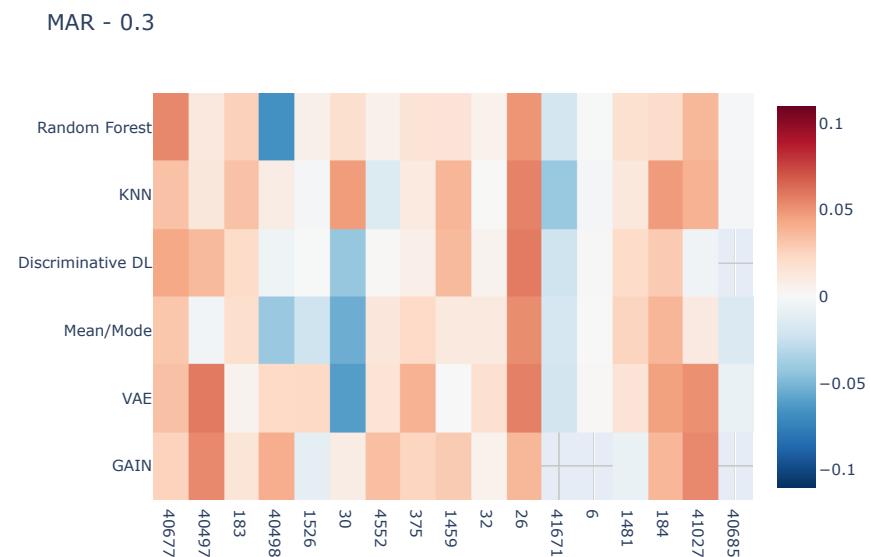


Abbildung I.7.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.3 Missing Fraction

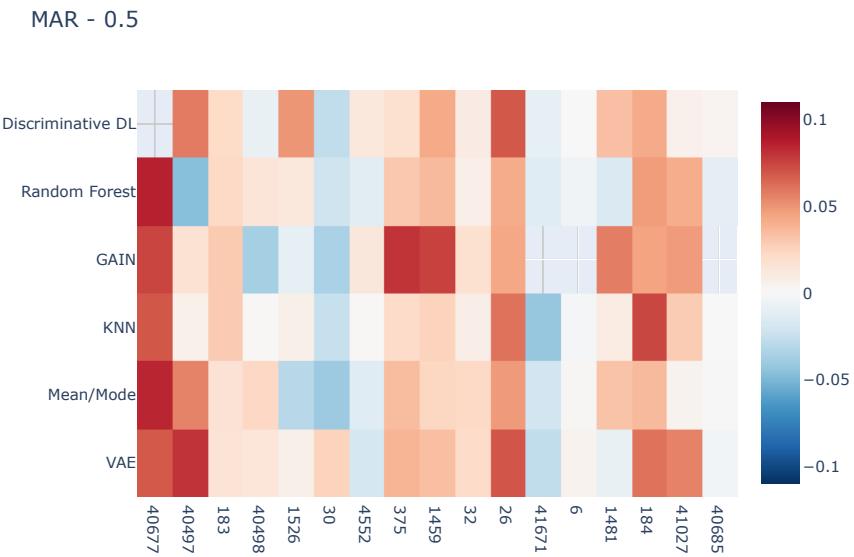


Abbildung I.8.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.5 Missing Fraction

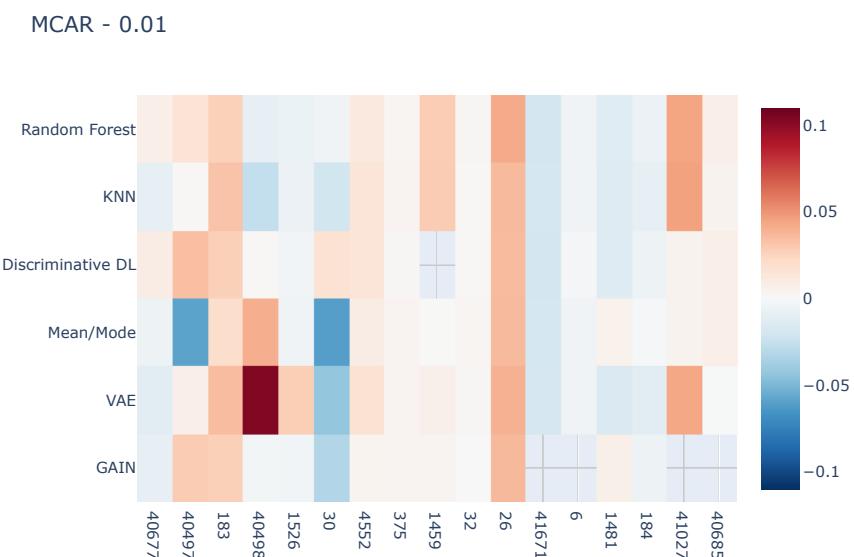


Abbildung I.9.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.01 Missing Fraction

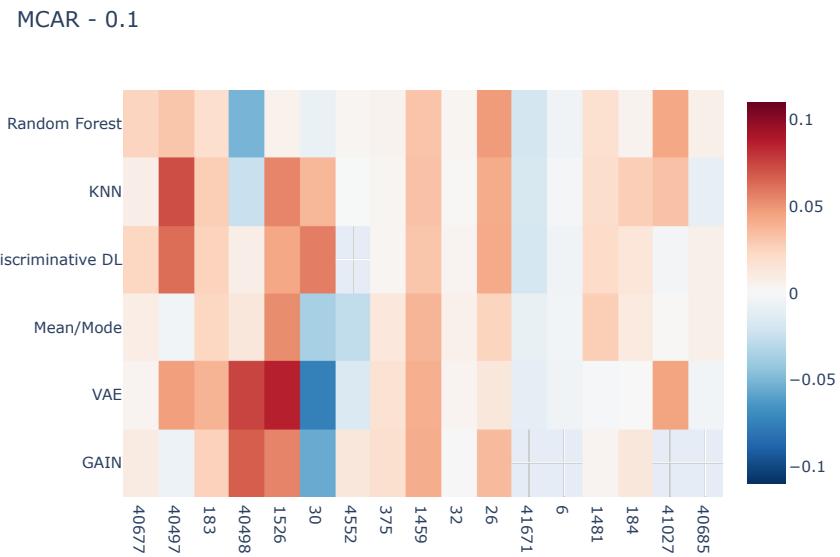


Abbildung I.10.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.1 Missing Fraction

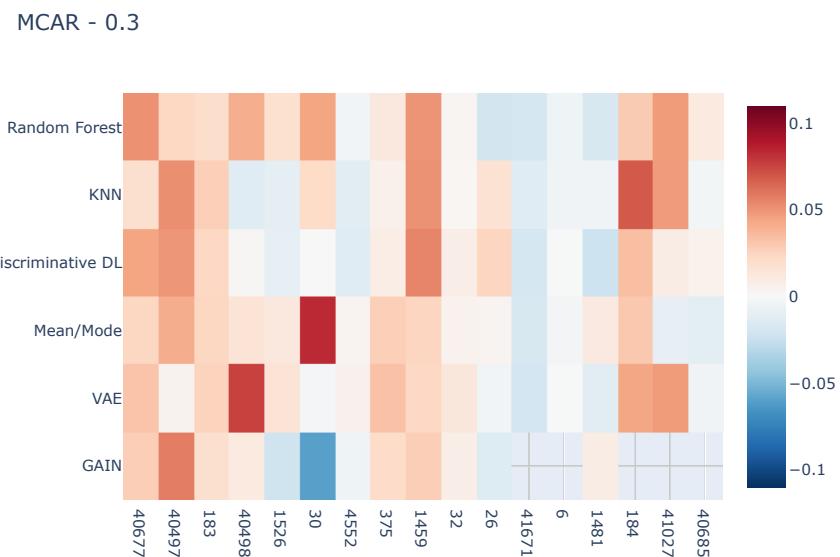


Abbildung I.11.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.3 Missing Fraction

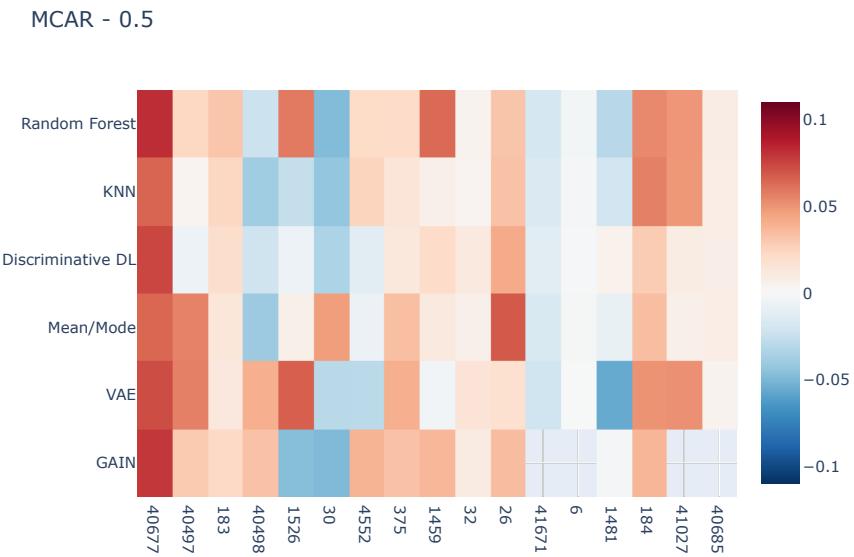


Abbildung I.12.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.5 Missing Fraction

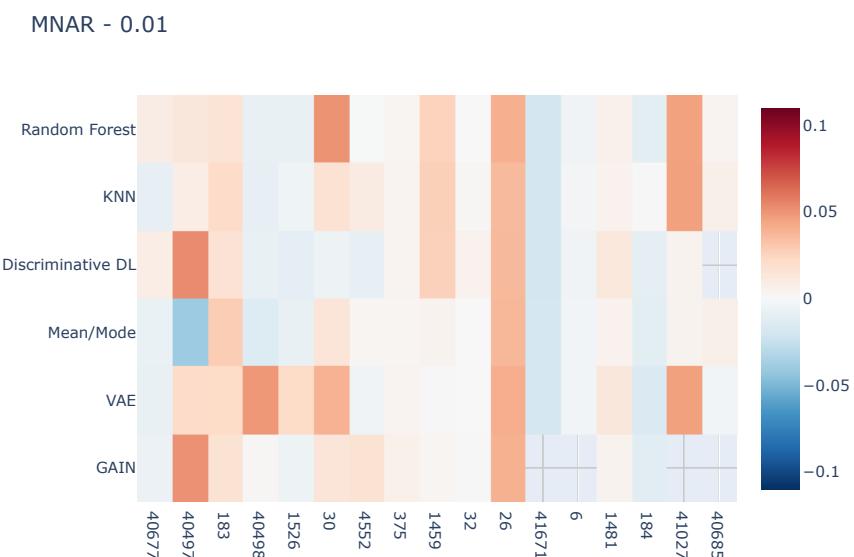


Abbildung I.13.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.01 Missing Fraction

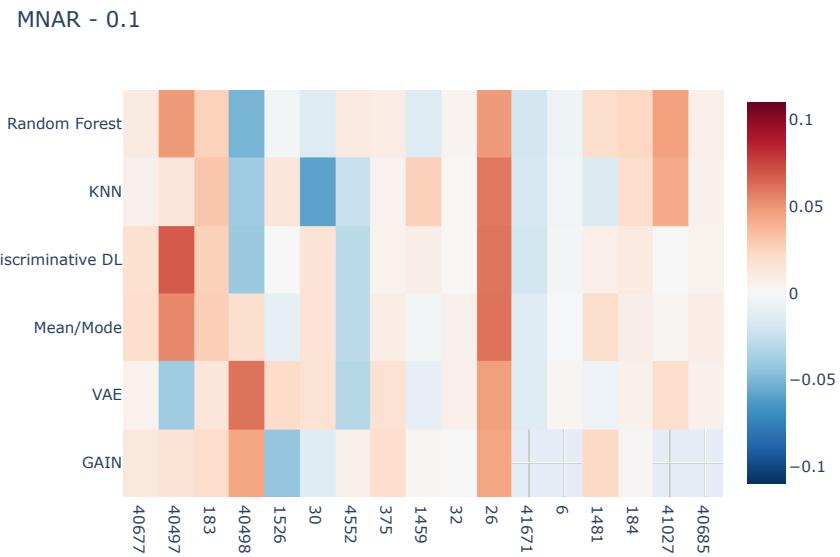


Abbildung I.14.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.1 Missing Fraction

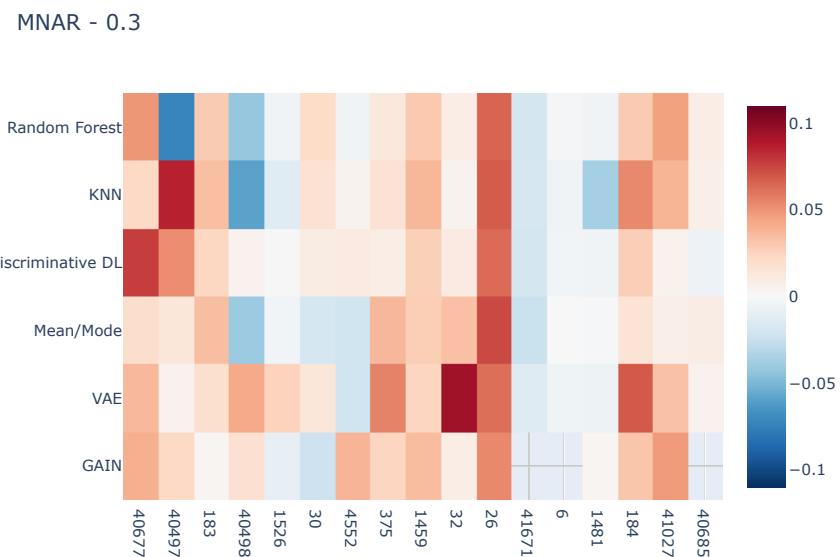


Abbildung I.15.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.3 Missing Fraction

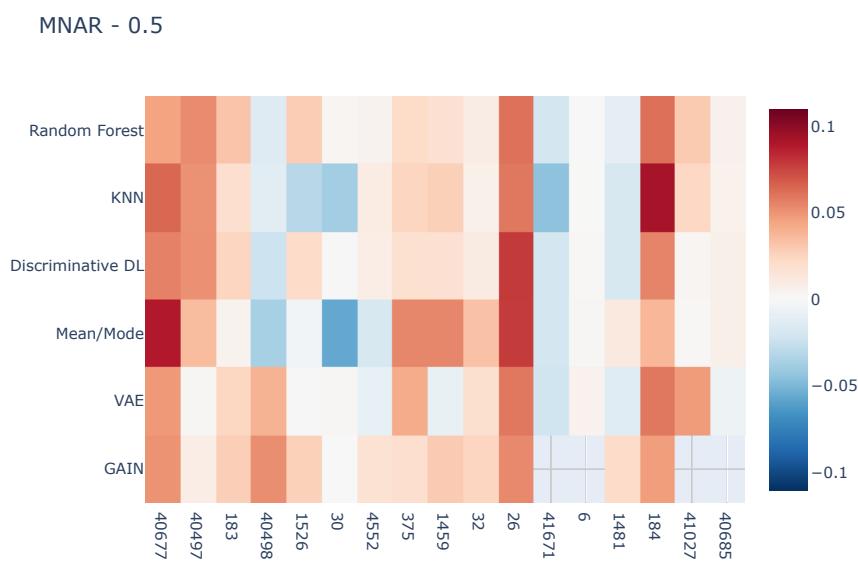


Abbildung I.16.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.5 Missing Fraction

J. Appendix Multiclass Classification Imputed - Subset Comparisons

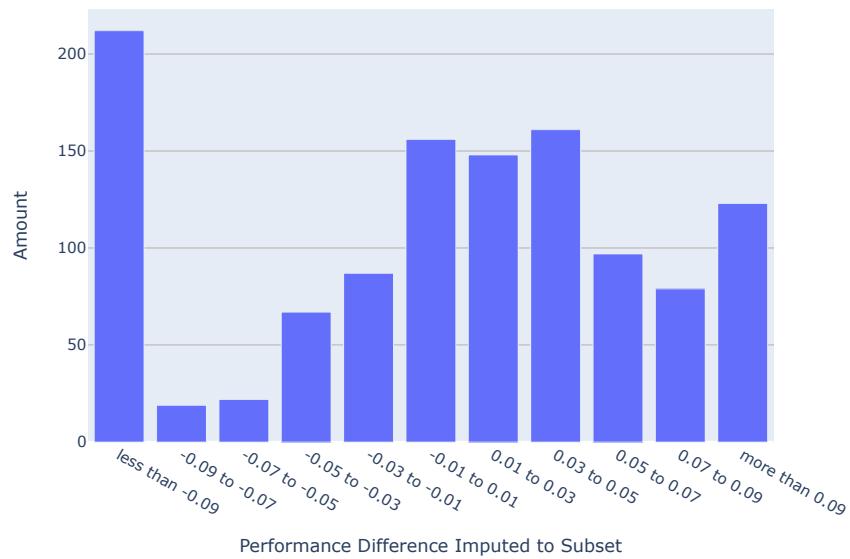


Abbildung J.1.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data

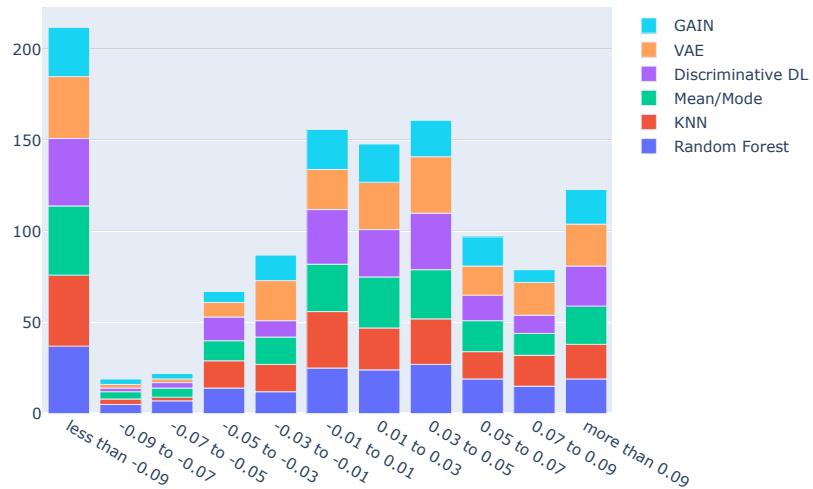


Abbildung J.2.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Imputation Method

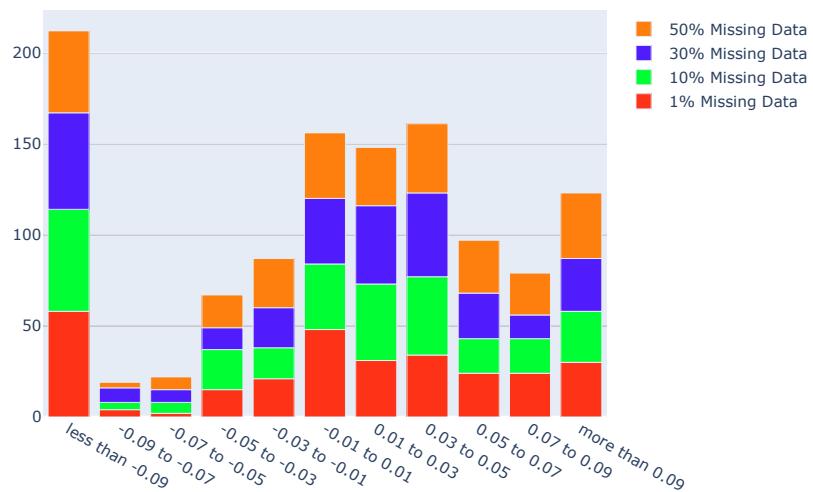


Abbildung J.3.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Fraction

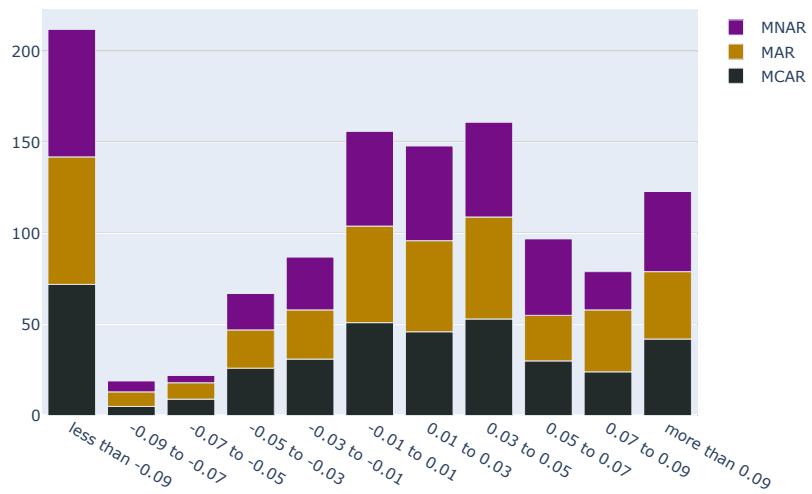


Abbildung J.4.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Pattern

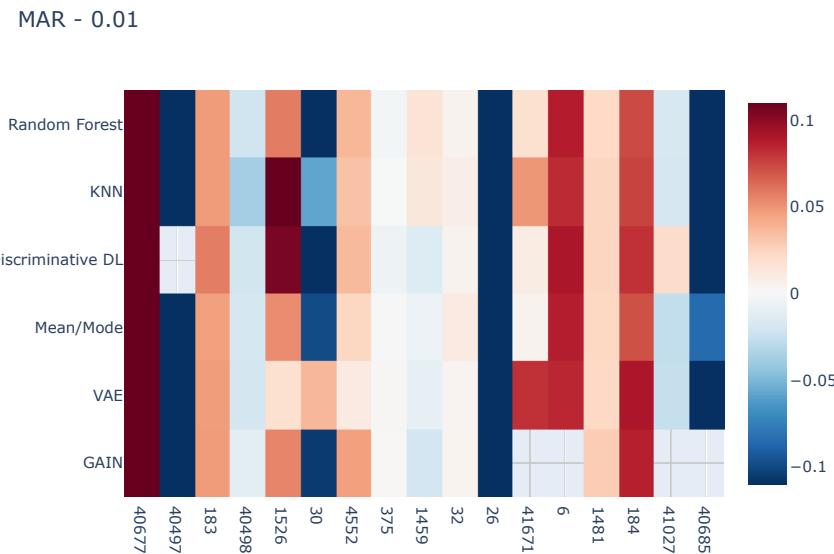


Abbildung J.5.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.01 Missing Fraction

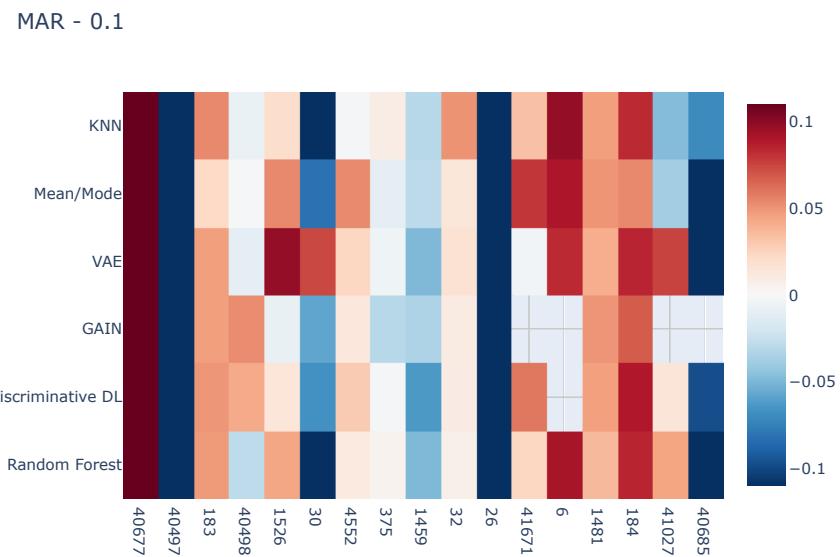


Abbildung J.6.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.1 Missing Fraction

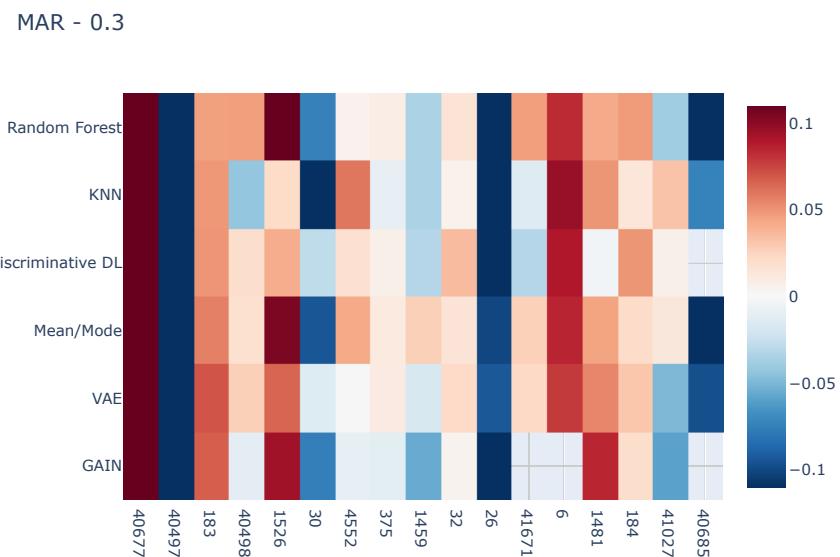


Abbildung J.7.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.3 Missing Fraction

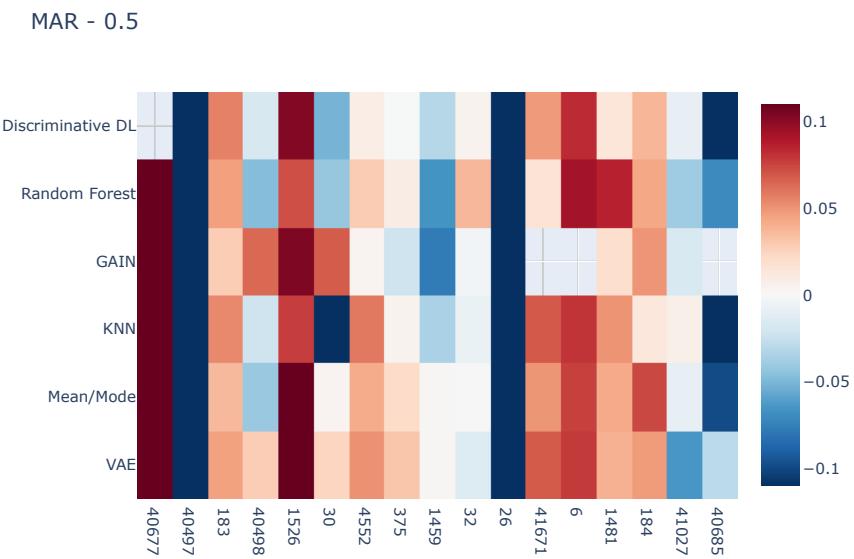


Abbildung J.8.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.5 Missing Fraction

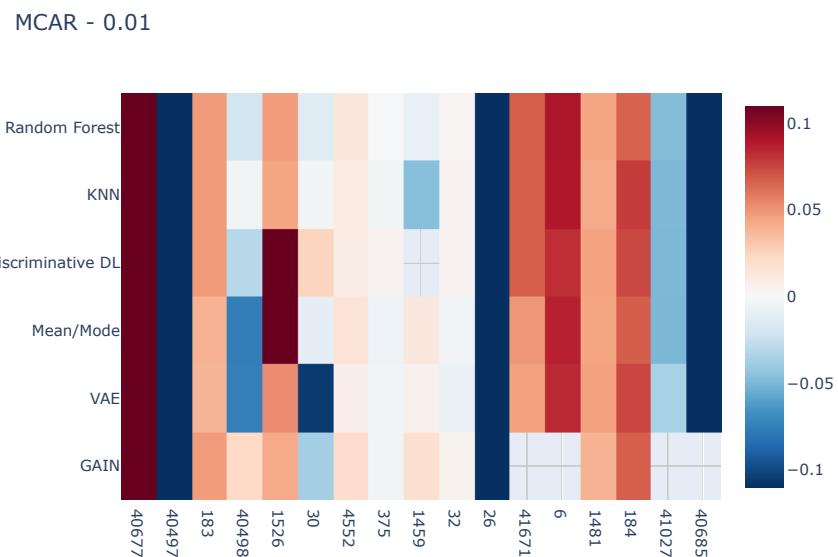


Abbildung J.9.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.01 Missing Fraction

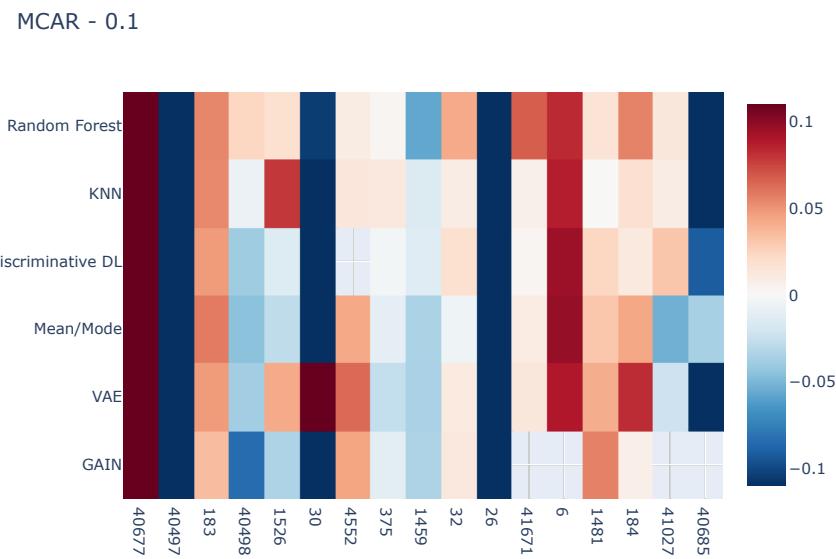


Abbildung J.10.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.1 Missing Fraction

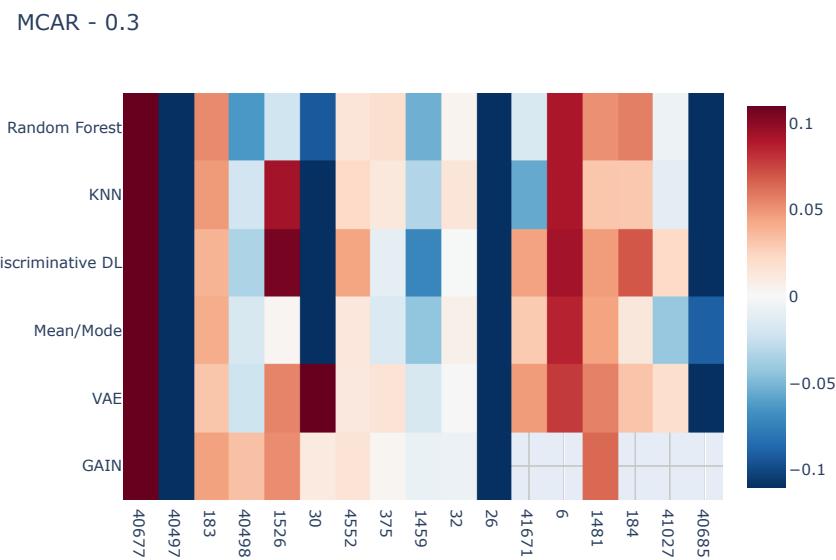


Abbildung J.11.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.3 Missing Fraction

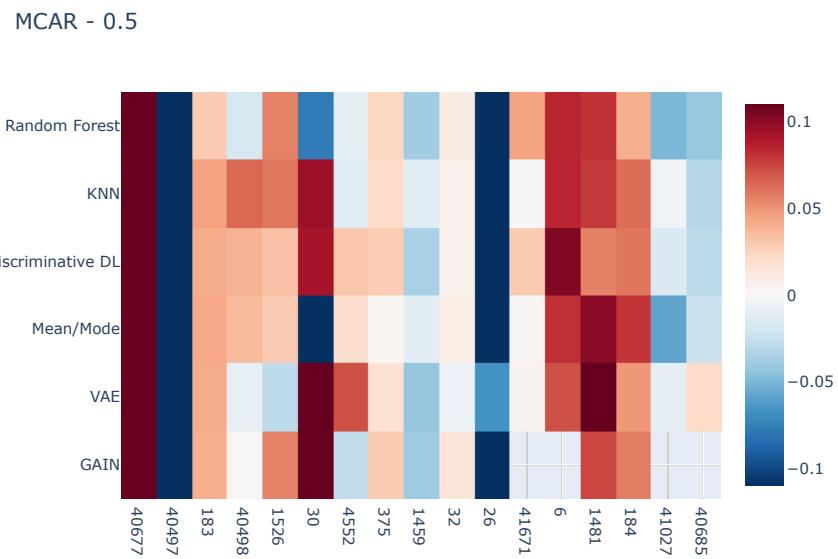


Abbildung J.12.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.5 Missing Fraction

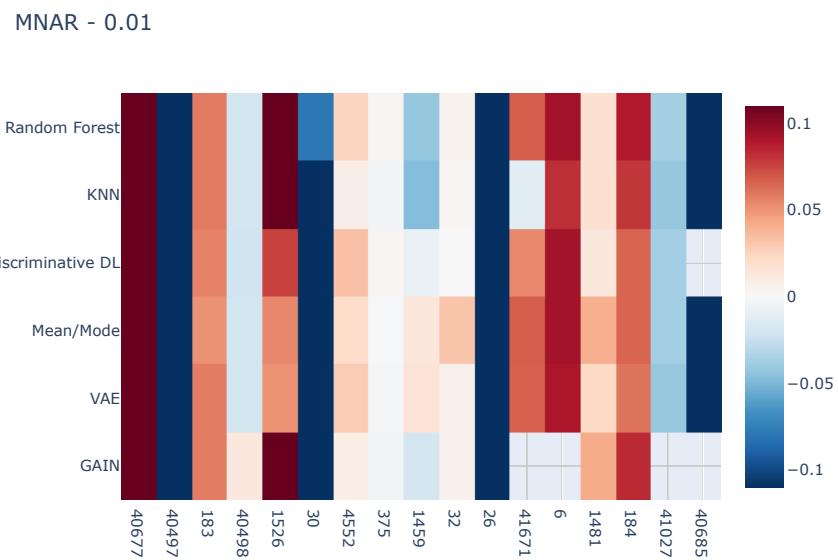


Abbildung J.13.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.01 Missing Fraction

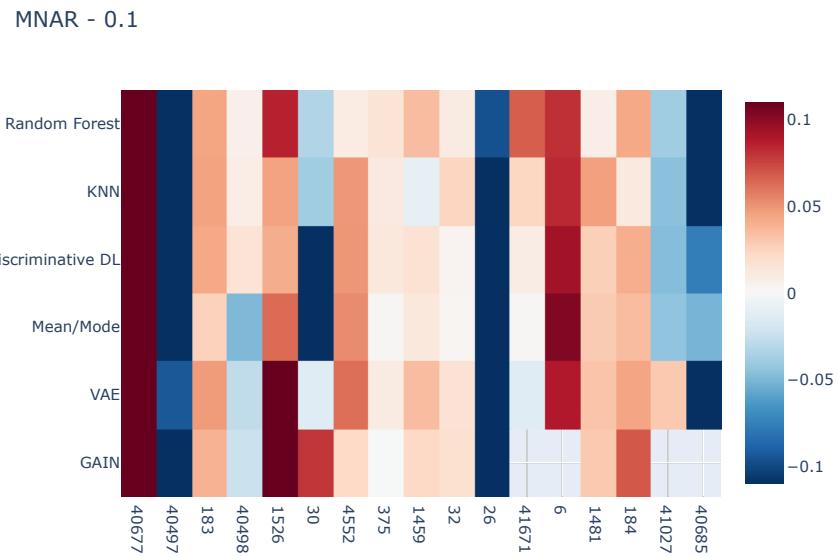


Abbildung J.14.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.1 Missing Fraction

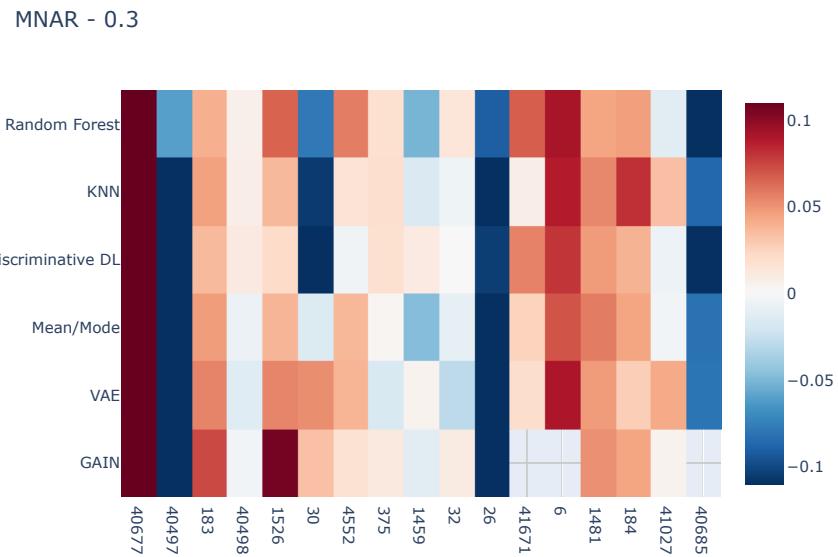


Abbildung J.15.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.3 Missing Fraction

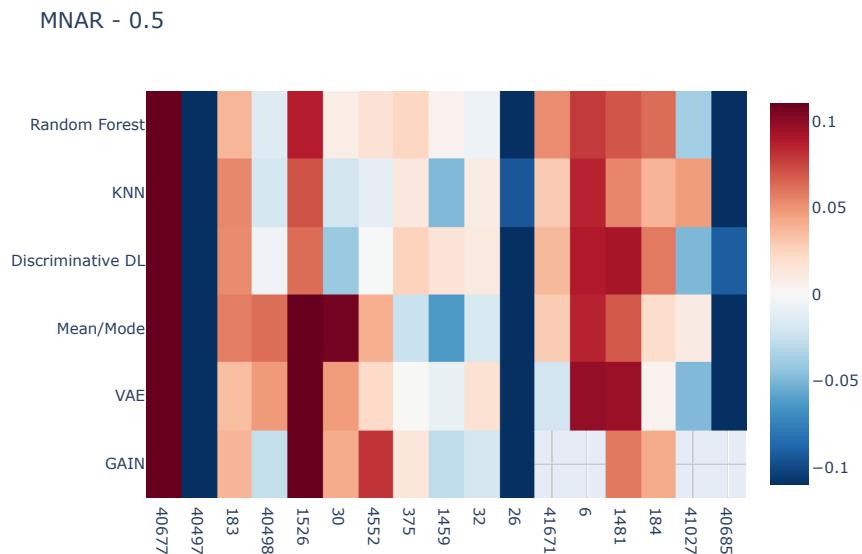


Abbildung J.16.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.5 Missing Fraction

K. Appendix Regression Baseline - Imputed Comparisons

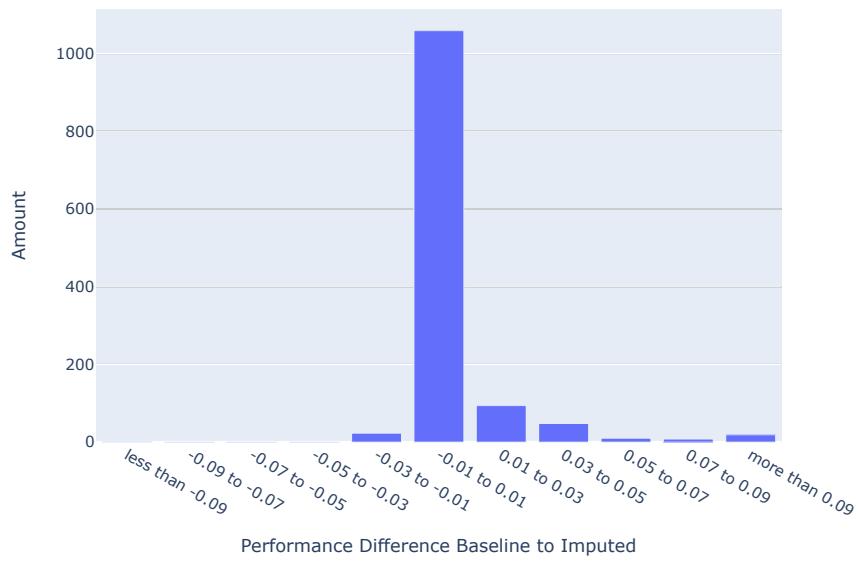


Abbildung K.1.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data

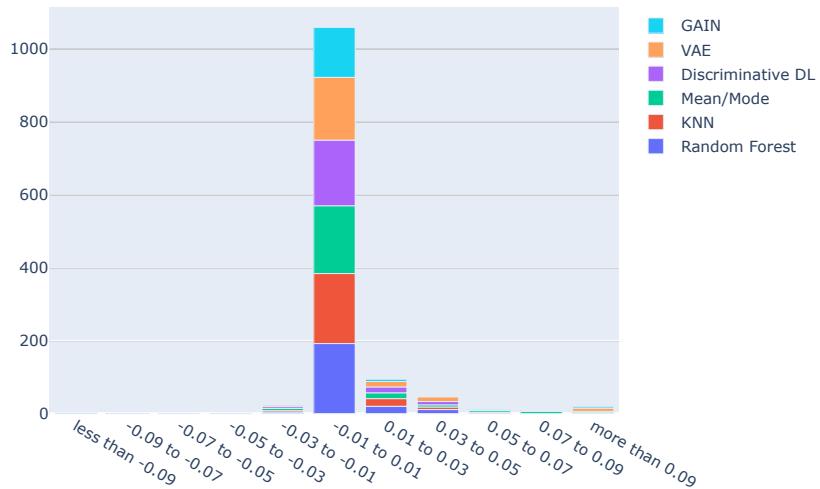


Abbildung K.2.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Imputation Method

Improvement Relative to Average Best for All Dataconstellations and Imputati

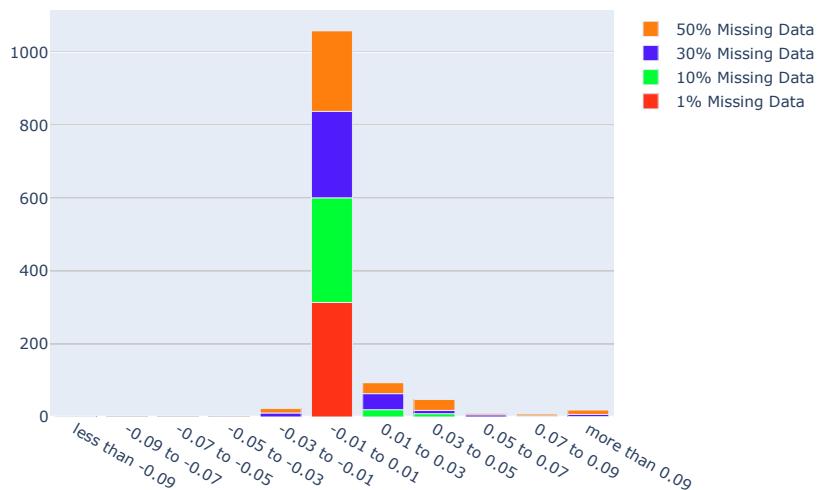


Abbildung K.3.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Fraction

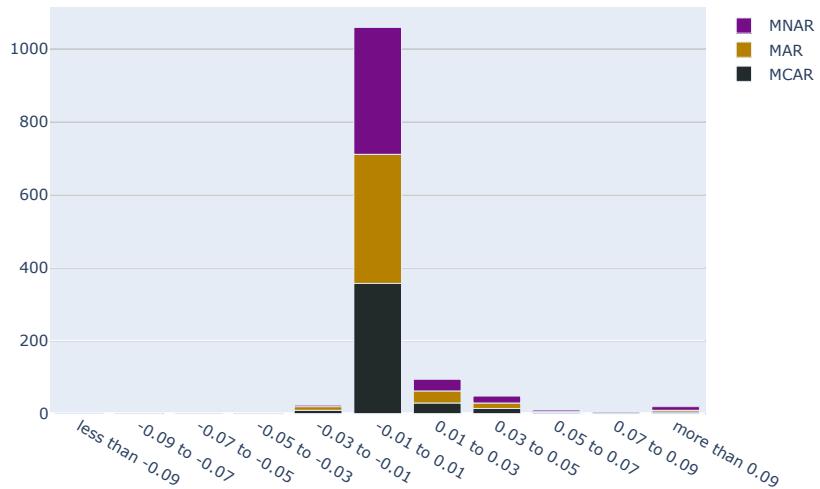


Abbildung K.4.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data by Missing Pattern

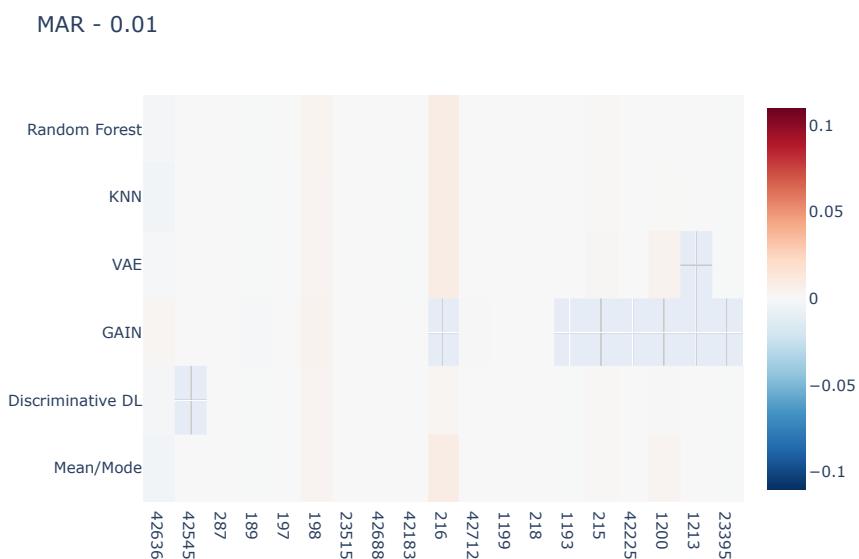


Abbildung K.5.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.01 Missing Fraction

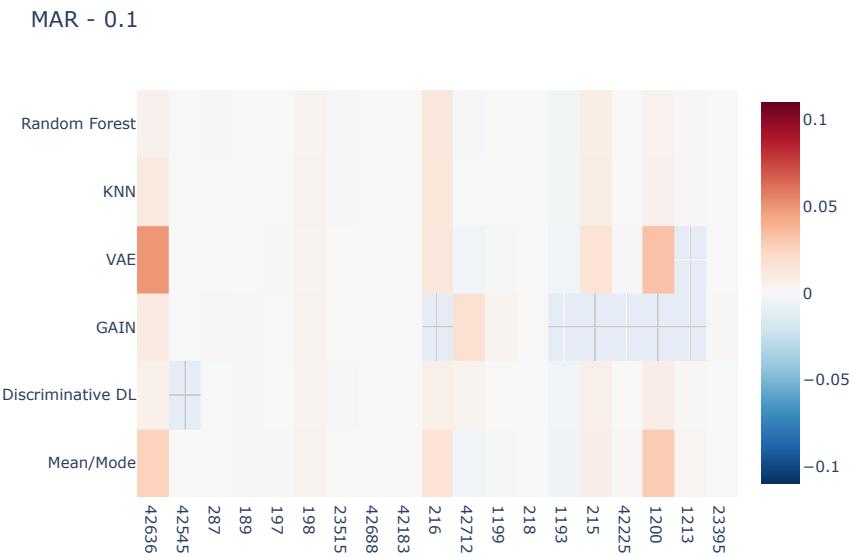


Abbildung K.6.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.1 Missing Fraction

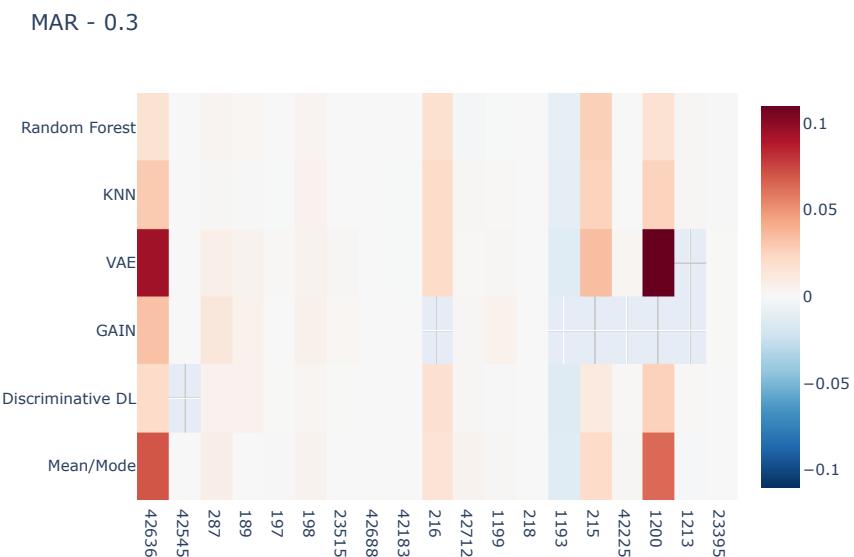


Abbildung K.7.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.3 Missing Fraction

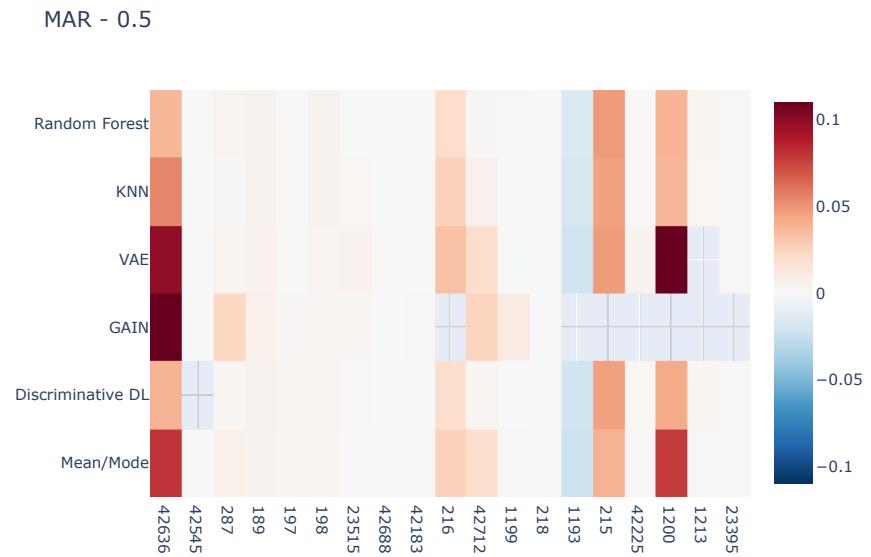


Abbildung K.8.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MAR and 0.5 Missing Fraction

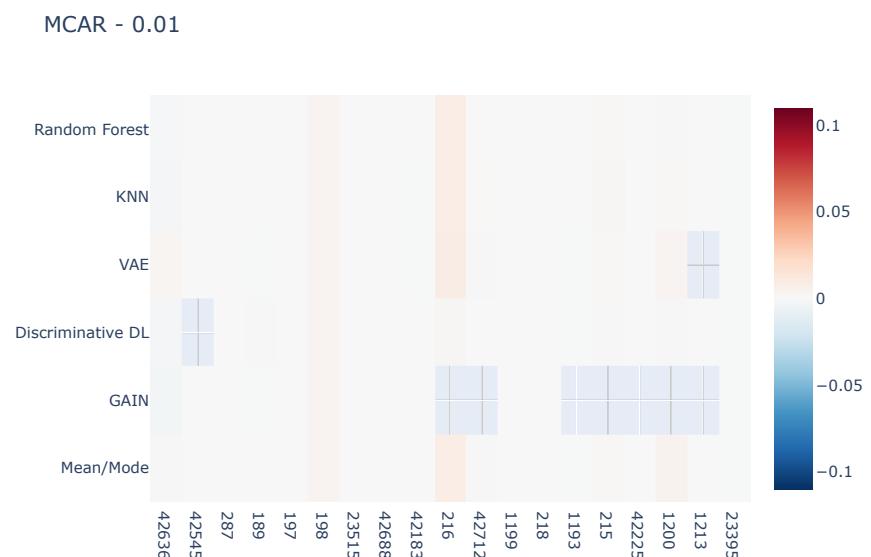


Abbildung K.9.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.01 Missing Fraction

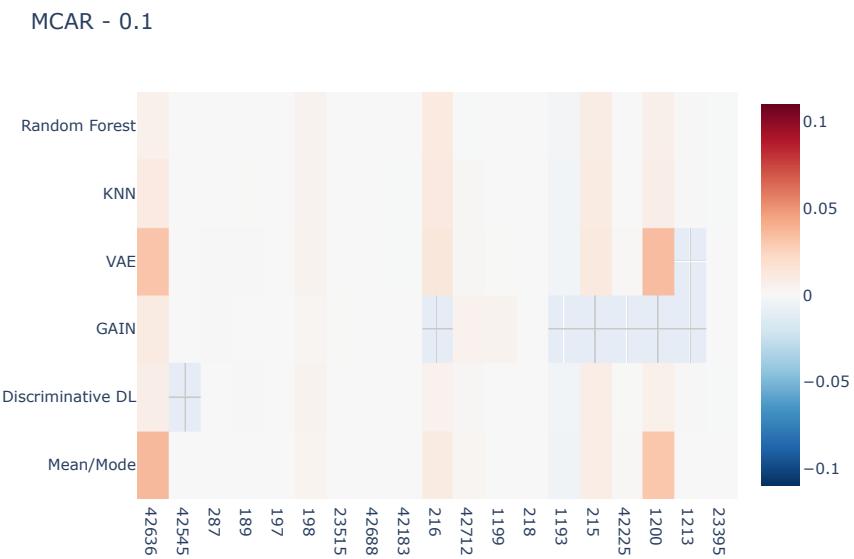


Abbildung K.10.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.1 Missing Fraction

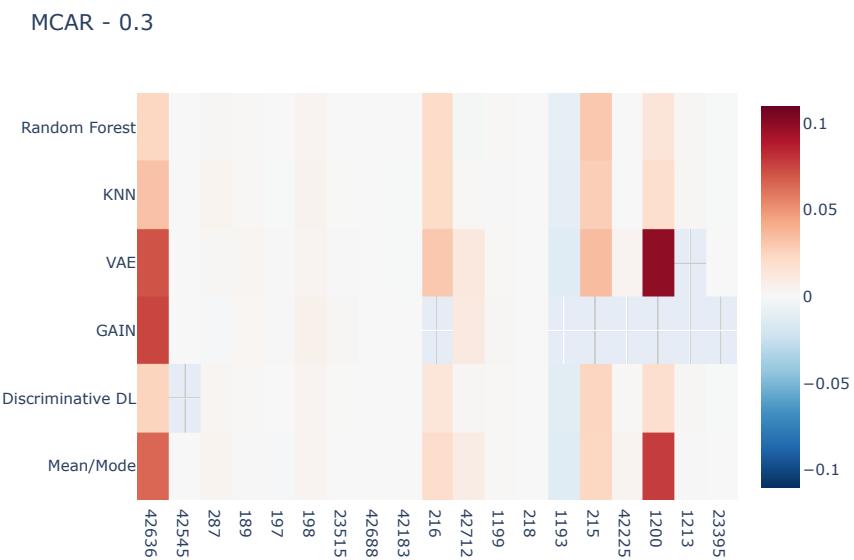


Abbildung K.11.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.3 Missing Fraction

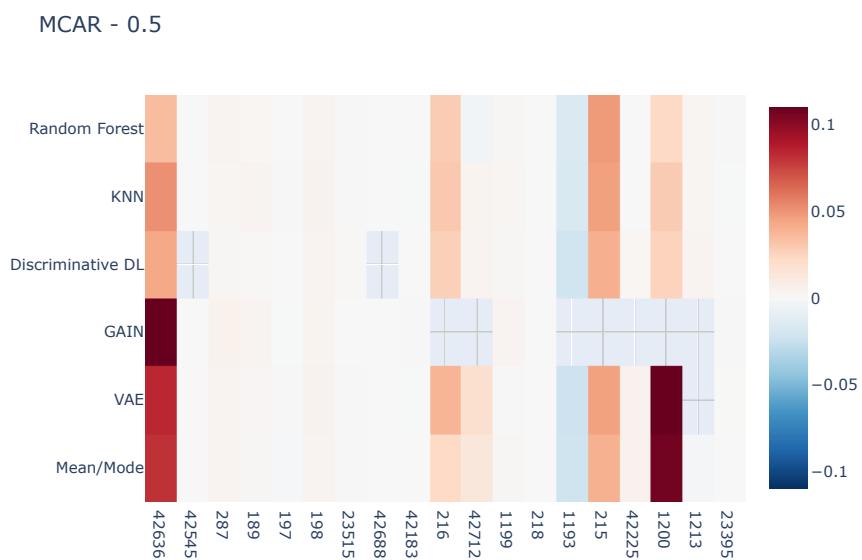


Abbildung K.12.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MCAR and 0.5 Missing Fraction

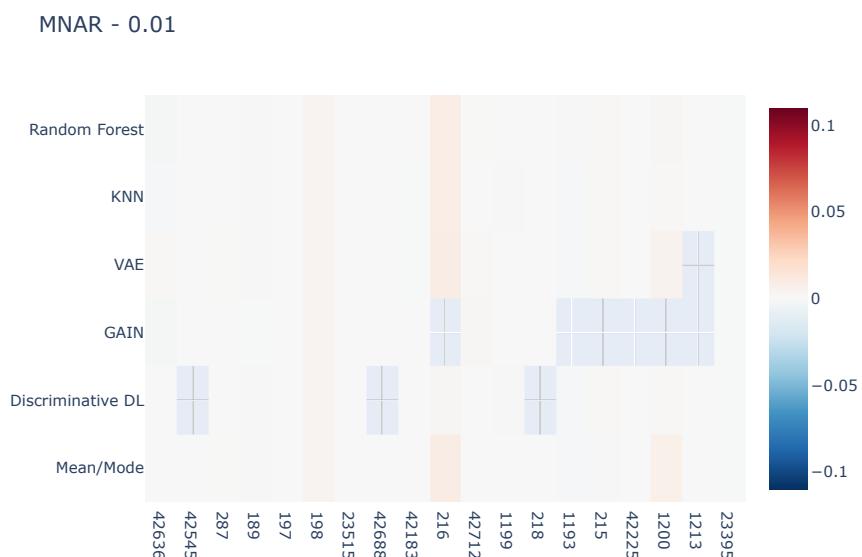


Abbildung K.13.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.01 Missing Fraction

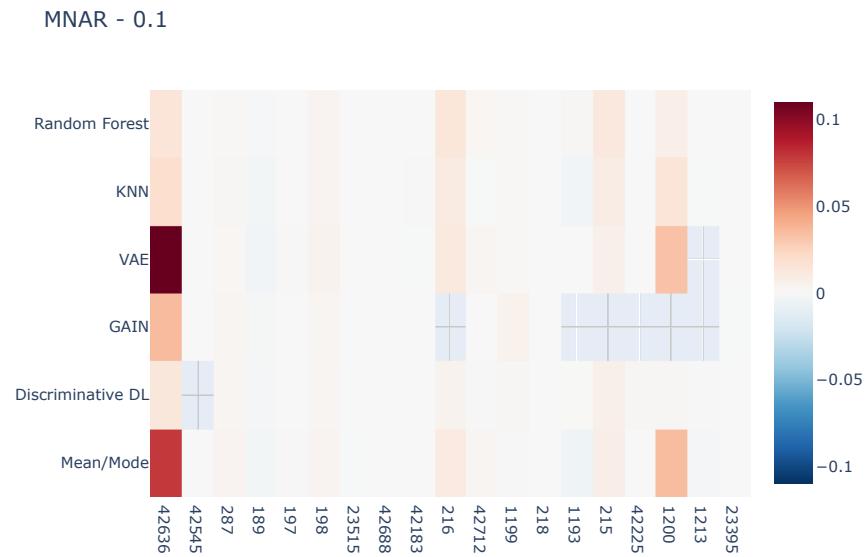


Abbildung K.14.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.1 Missing Fraction

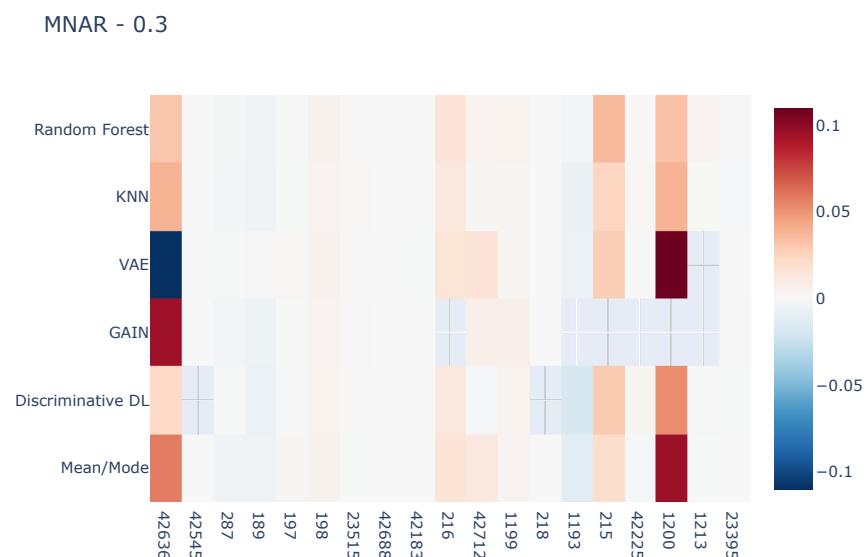


Abbildung K.15.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.3 Missing Fraction

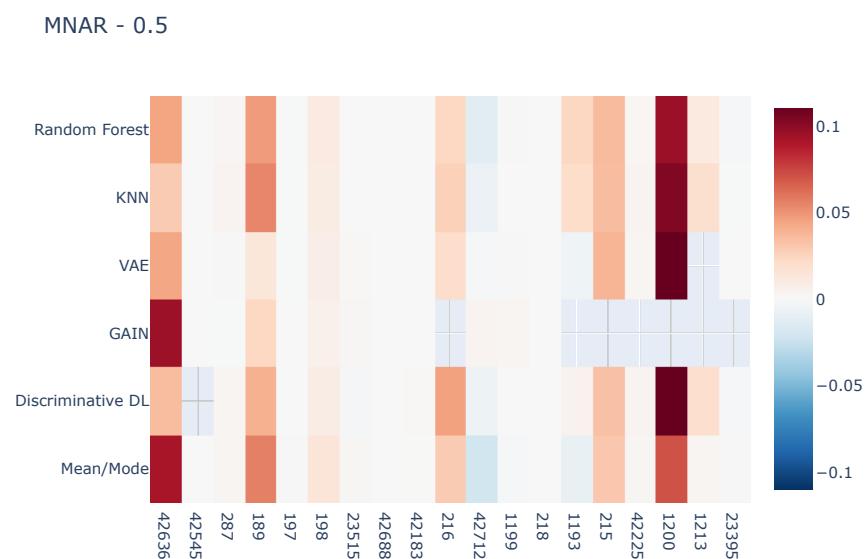


Abbildung K.16.: Improvements for the Baseline Model Relative to the Model Trained on Imputed Data per Dataset for MNAR and 0.5 Missing Fraction

L. Appendix Regression Imputed - Subset Comparisons

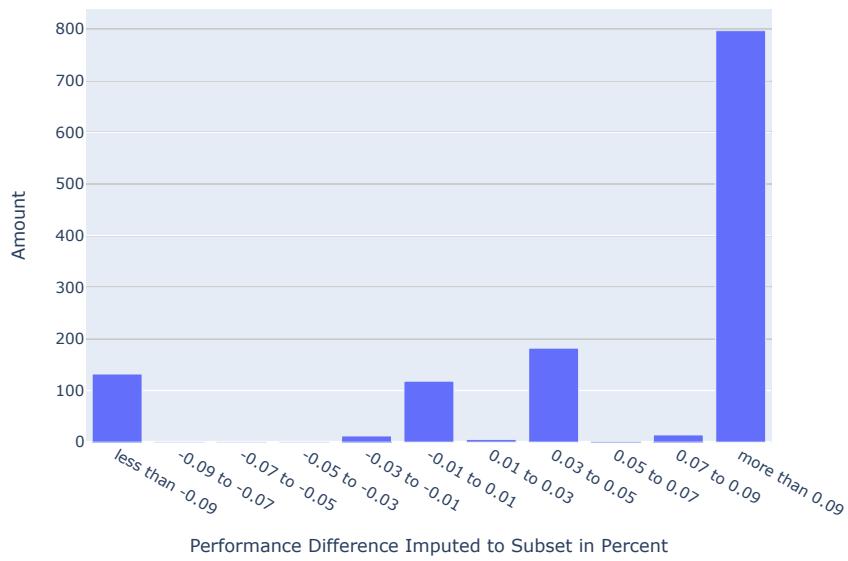


Abbildung L.1.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data

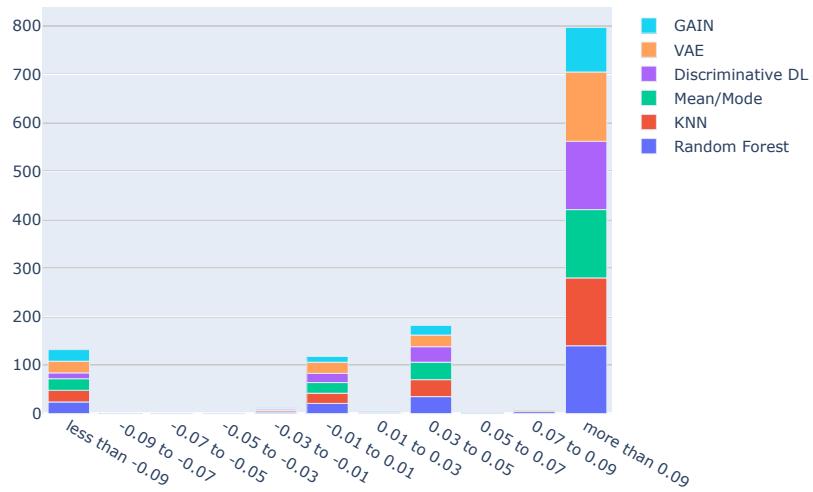


Abbildung L.2.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Imputation Method

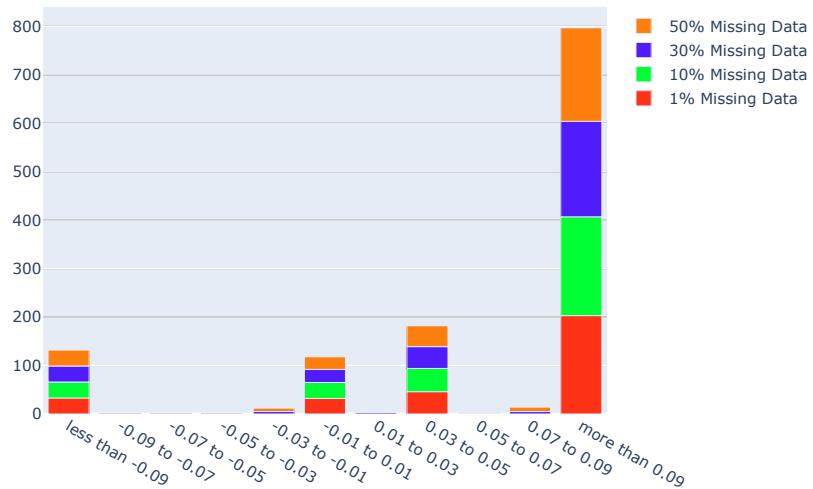


Abbildung L.3.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Fraction

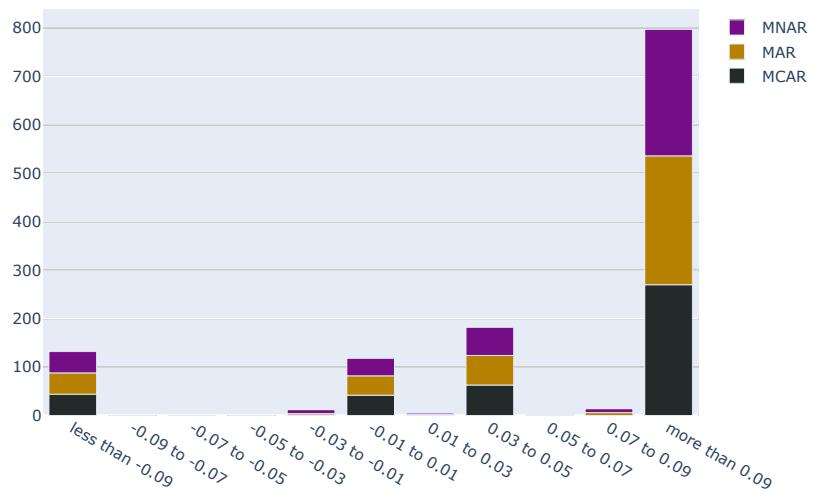


Abbildung L.4.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data by Missing Pattern

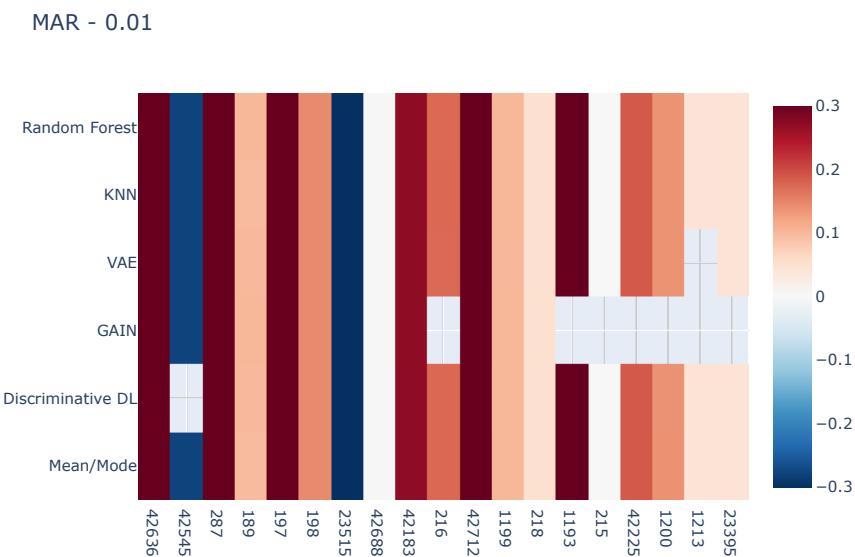


Abbildung L.5.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.01 Missing Fraction

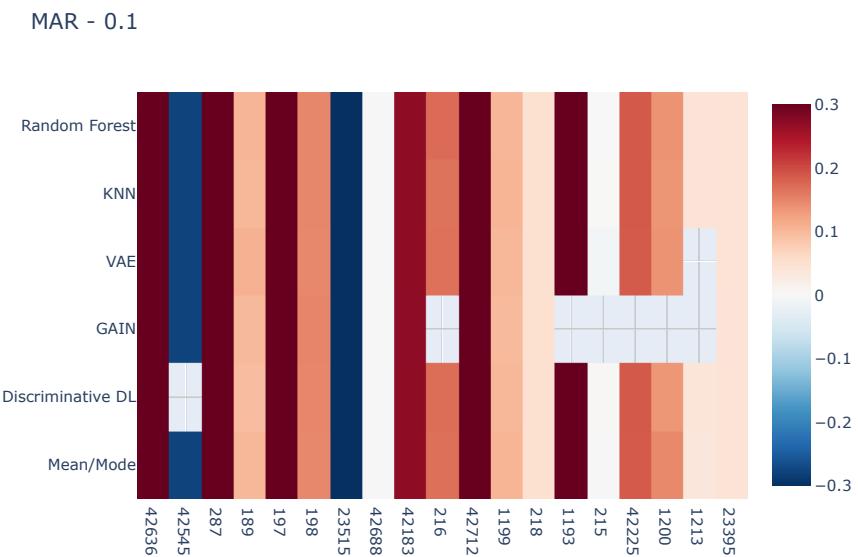


Abbildung L.6.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.1 Missing Fraction

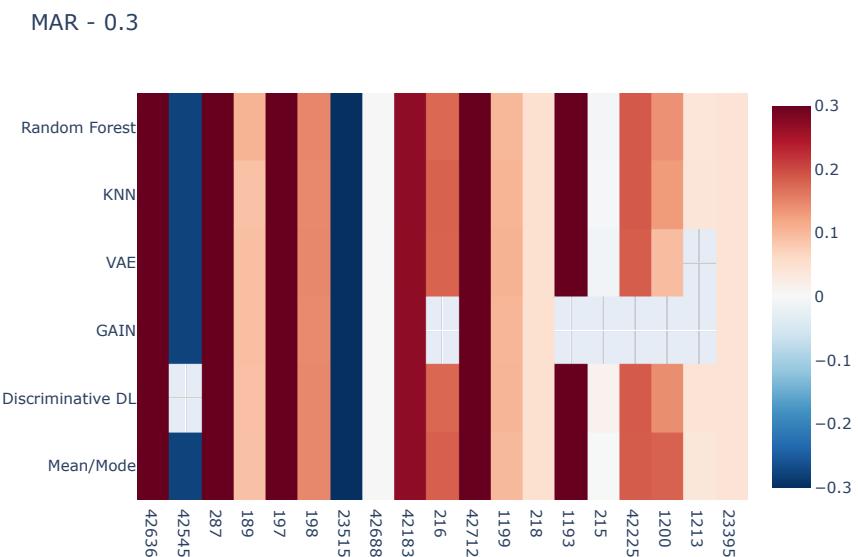


Abbildung L.7.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.3 Missing Fraction

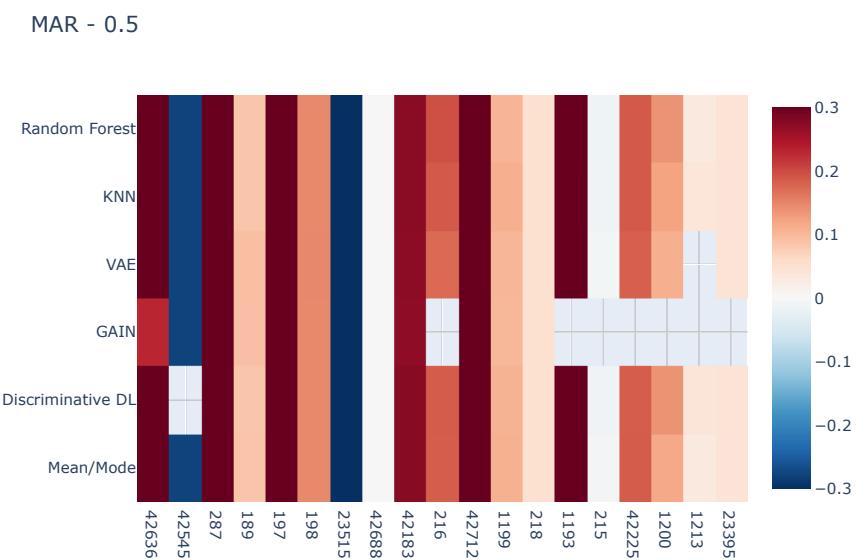


Abbildung L.8.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MAR and 0.5 Missing Fraction

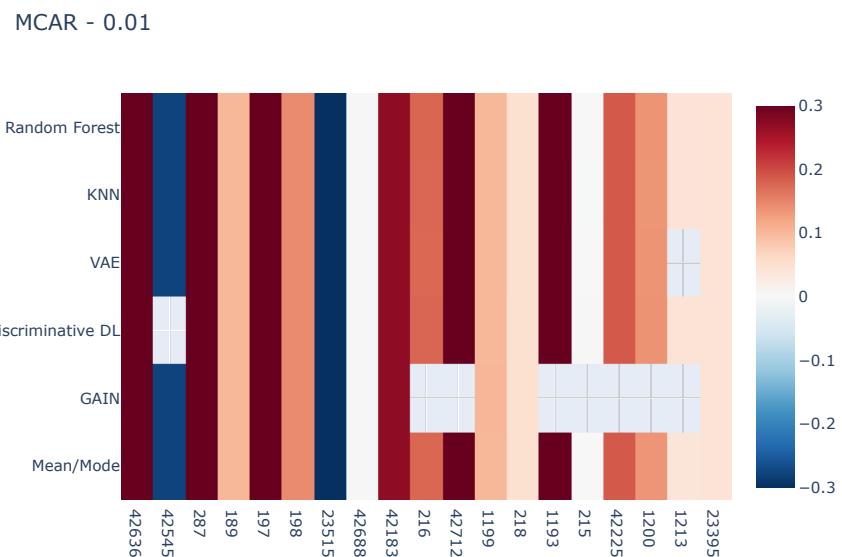


Abbildung L.9.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.01 Missing Fraction

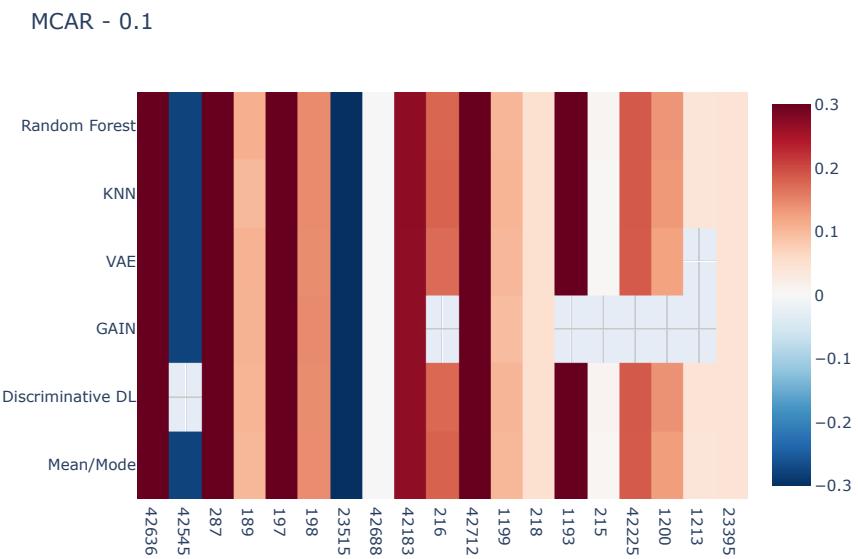


Abbildung L.10.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.1 Missing Fraction

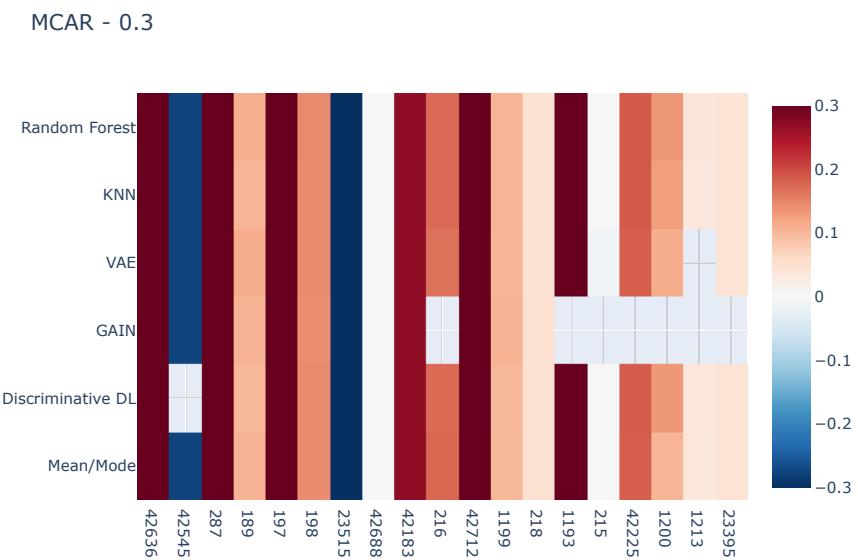


Abbildung L.11.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.3 Missing Fraction

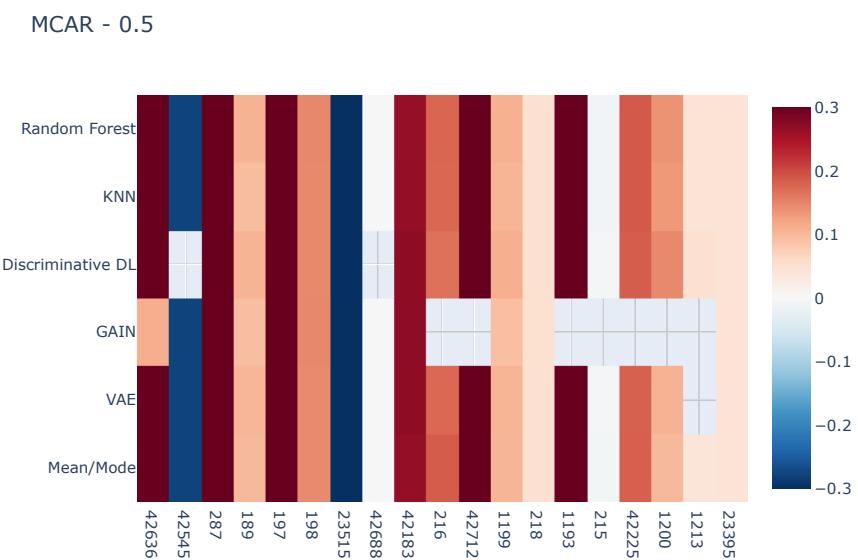


Abbildung L.12.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MCAR and 0.5 Missing Fraction

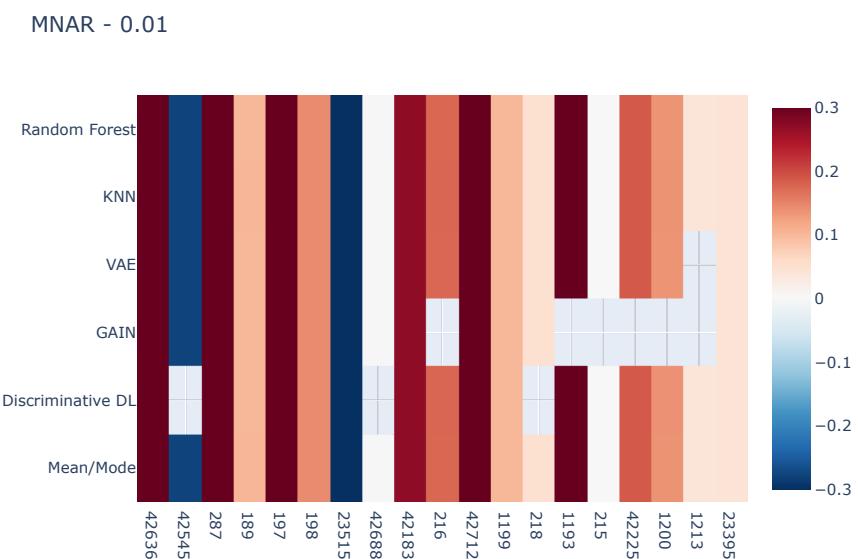


Abbildung L.13.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.01 Missing Fraction

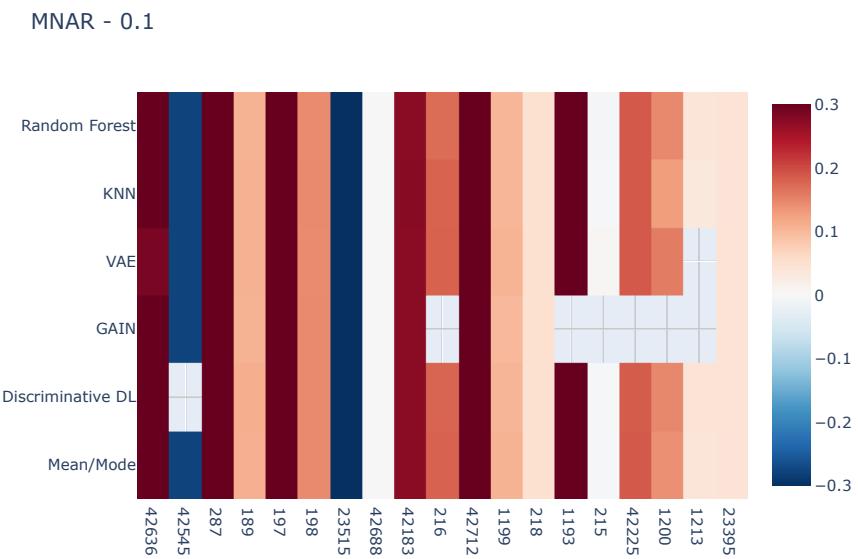


Abbildung L.14.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.1 Missing Fraction

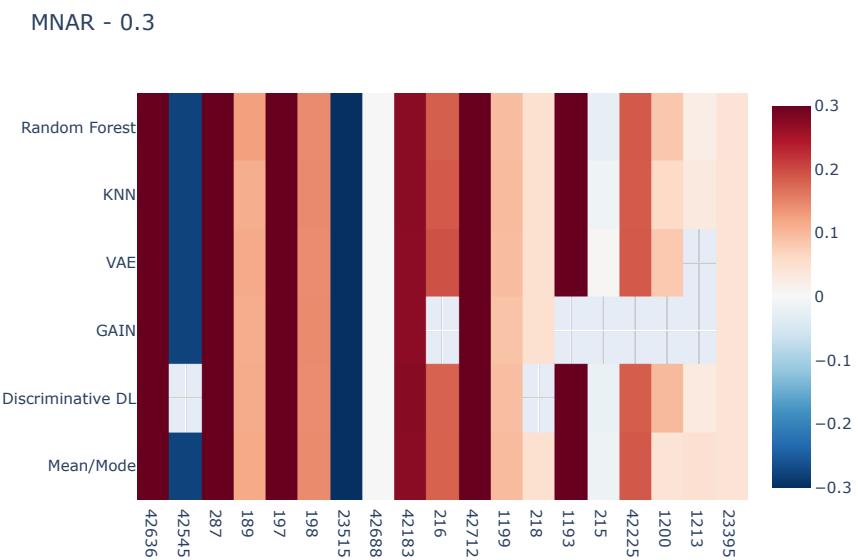


Abbildung L.15.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.3 Missing Fraction

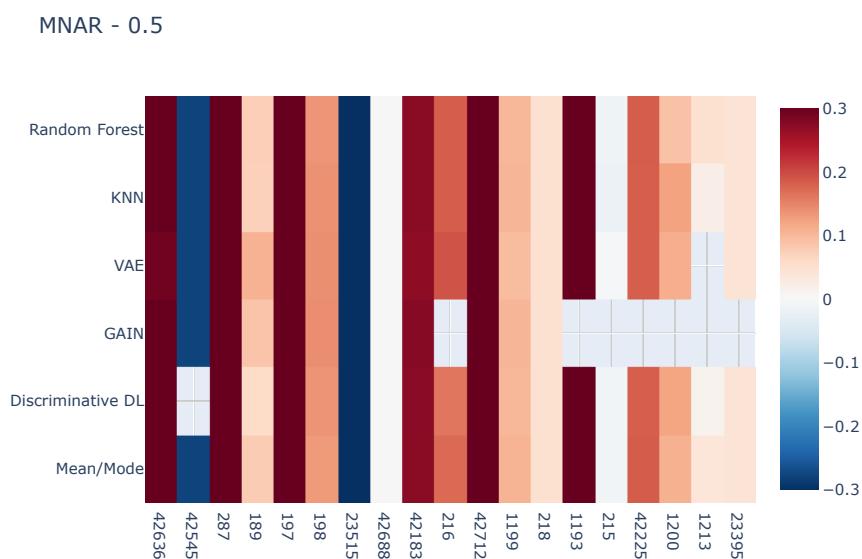


Abbildung L.16.: Improvements for the Model Trained on Imputed Data Relative to the Model Trained on Subset Data per Dataset for MNAR and 0.5 Missing Fraction

M. Important Changes in the Framework

Change File Path: src/jenga/src/jenga/basis.py

Change Location: class Task / def fit_baseline_model

Preparation for changed model training (with imputed data) by skipping the previously implemented *SimpleImputer*.

```
1      categorical_preprocessing = Pipeline(
2          [
3              #('mark_missing',
4              #    SimpleImputer(strategy='most_frequent'))),
5              ('one_hot_encode',
6                  OneHotEncoder(handle_unknown='ignore'))
7          ]
8      )
9
10     numerical_preprocessing = Pipeline(
11         [
12             #('mark_missing', SimpleImputer(strategy='mean')),
13             ('scaling', StandardScaler())
14         ]
15     )
```

Quelltext M.1: Preparation for Model Training with Imputed Data

Change File Path: src/jenga/src/jenga/basis.py

Change Location: BinaryClassificationTask & MultiClassClassification & RegressionTask / def _get_pipeline_grid_scorer_tuple

To ensure reproducibility the random seed is set for *SGDClassifier* and *SGDRegressor*.

```
1     rng = np.random.RandomState(5)
2     print(rng, "random seed for SGD Classifier")
3     pipeline = Pipeline(
4         [
5             ('features', feature_transformation),
6             ('learner', SGDClassifier(max_iter=1000,
7                                         n_jobs=-1, random_state=rng))
8         ]
)
```

Quelltext M.2: SGDClassifier Random Seed

```
1     rng = np.random.RandomState(5)
2     print(rng, "random seed for SGD #regressor")
3     pipeline = Pipeline(
4         [
5             ('features', feature_transformation),
6             ('learner', SGDRegressor(max_iter=1000,
7                                      random_state=rng))
8         ]
)
```

Quelltext M.3: SGDRegressor Random Seed

Change File Path: src/jenga/src/jenga/basis.py

Change Location: class TabularCorruption / def sample_rows

Changed the approach to how the dependent column for MAR is picked. The original code rearranges the items randomly every time, that is bypassing the static set random seed.

```
1 elif self.sampling.endswith('AR'):
2     columns_temp_ar = data.loc[:, data.columns != self.column]
3     list_for_columns = list(columns_temp_ar.columns)
4     depends_on_col = np.random.choice(list_for_columns)
5     print(depends_on_col)
6     rows =
        data[depends_on_col].sort_values().iloc[perc_idx].index
```

Quelltext M.4: Column Selection MAR Experiments

Change File Path: src/jenga/src/jenga/tasks/openml.py

Change Location: class OpenMLTask

Set static random seed for training/test split to ensure that experiments are deterministic.

```
1  class OpenMLTask(Task):
2
3      def __init__(self, openml_id: int, train_size: float = 0.8,
4                   seed: Optional[int] = None):
5          seed=32
6          X, y = fetch_openml(data_id=openml_id, as_frame=True,
7                               return_X_y=True, cache=False)
8          train_data, test_data, train_labels, test_labels =
9              train_test_split(X, y, train_size=train_size)
10         categorical_columns = [
11             column for column in X.columns
12             if pd.api.types.is_categorical_dtype(X[column])
13         ]
14         numerical_columns = [
15             column for column in X.columns
16             if pd.api.types.is_numeric_dtype(X[column]) and
17                 column not in categorical_columns
18         ]
19
20         super().__init__(
21             train_data=train_data,
22             train_labels=train_labels,
23             test_data=test_data,
24             test_labels=test_labels,
25             categorical_columns=categorical_columns,
26             numerical_columns=numerical_columns,
27             is_image_data=False,
28             seed=seed
29         )
```

Quelltext M.5: OpenML Random Seed

Change File Path: src/jenga/src/jenga/tasks/openml.py

Change Location: class OpenMLTask

Adjustment for Subset Experiment, which selects randomly 10% of the dataset for further experiments. The random seed is set to ensure that the experiment remains deterministic.

```
1  class OpenMLTask(Task):
2
3      def __init__(self, openml_id: int, train_size: float = 0.8,
4                   seed: Optional[int] = None):
5
6          X, y = fetch_openml(data_id=openml_id, as_frame=True,
7                               return_X_y=True, cache=False)
8
9          before_subset = pd.concat([X, y], axis=1)
10
11         np.random.seed(32)
12         len_dataset = len(before_subset)
13         print(len_dataset,'Length of dataset')
14
15         df_subset = before_subset.sample(frac=0.10)
16         y_df = df_subset.iloc[:, -1:]
17         y = y_df.iloc[:, 0]
18         X = df_subset[df_subset.columns[:-1]]
19
20         seed=32
21         train_data, test_data, train_labels, test_labels =
22             train_test_split(X, y, train_size=train_size)
23         categorical_columns = [
24             column for column in X.columns
25             if pd.api.types.is_categorical_dtype(X[column])
26         ]
27         numerical_columns = [
28             column for column in X.columns
29             if pd.api.types.is_numeric_dtype(X[column]) and
30                 column not in categorical_columns
31         ]
```

Quelltext M.6: Subset Experiment Adjustments

Change File Path: src/data_imputation_paper/experiment.py

Change Location: class Experiment / def run

The random seed is set to make experiments and multiple runs deterministic.

```

1  for task_id, task_class in self._task_id_class_tuples:
2      self._result[task_id] = {}
3
4      task = task_class(openml_id=task_id)
5
6      for missing_type in self._missing_types:
7          self._result[task_id][missing_type] = {}
8
9      for missing_fraction in self._missing_fractions:
10         self._result[task_id][missing_type][missing_fraction]
11            = {}
12
13     for strategy in self._strategies:
14
15         experiment_path = self._base_path / f"{task_id}"
16             / missing_type / f"{missing_fraction}" /
17             f"{strategy}"
18
19     try:
20         evaluator =
21             self.strategy_to_EvaluatorClass[strategy](
22                 task=task,
23                 missing_fraction=missing_fraction,
24                 missing_type=missing_type,
25                 target_column=target_column,
26                 imputer_class=self._imputer_class,
27                 imputer_args=self._imputer_arguments,
28                 path=experiment_path,
29                 seed=42
30             )
31         evaluator.evaluate(self._num_repetitions)
32         result = evaluator._result

```

Quelltext M.7: Random Seed for Multiple Repetitions of Experiments

Change File Path: `src/data_imputation_paper/evaluation_corrupted.py`

Change Location: Class Evaluator / def evaluate

Adjust function so that now the imputed training and test data is used, instead of corrupted training and test data.

```

1  train_data_corrupted, test_data_corrupted = self._discard_values(
2      task=self._task,
3      to_discard_columns=self._discard_in_columns,
4      missing_fraction=self._missing_fraction,
5      missing_type=self._missing_type,
6  )
7
8
9  imputer = self._imputer_class(**self._imputer_arguments)
10
11 start_time = time.time()
12 imputer.fit(train_data_corrupted.copy(), [target_column])
13 elapsed_time = time.time() - start_time
14
15 train_imputed, train_imputed_mask =
16     imputer.transform(train_data_corrupted)
16 test_imputed, test_imputed_mask =
17     imputer.transform(test_data_corrupted)
18
18 if result_temp._baseline_performance is None:
19     base_model =
20         self._task.fit_baseline_model(train_imputed.copy(),
21             self._task.train_labels)
20 self._task._baseline_model = base_model
21
22 predictions = self._task._baseline_model.predict(test_imputed)
23 result_temp._baseline_performance =
24     self._task.score_on_test_data(predictions)
25
25 # NOTE: masks are DataFrames => append expects Series
26 result_temp.append(
27     target_column=target_column,
28     train_data_imputed=train_imputed,
29     test_data_imputed=test_imputed,
30     test_data_corrupted=test_data_corrupted,
31     train_imputed_mask=train_imputed_mask[target_column],
32     test_imputed_mask=test_imputed_mask[target_column],
33     elapsed_time=elapsed_time,
34     best_hyperparameters=imputer.get_best_hyperparameters()
35 )

```

Quelltext M.8: Adjustment for the Usage of the Imputed Training and Test Data