Data science & machine learning homework

Production Alignment Data Prediction

By Tomáš Čajan

Contents of this presentation

1 – Task Description

2 – EDA

3 – Machine Learning models and their performance

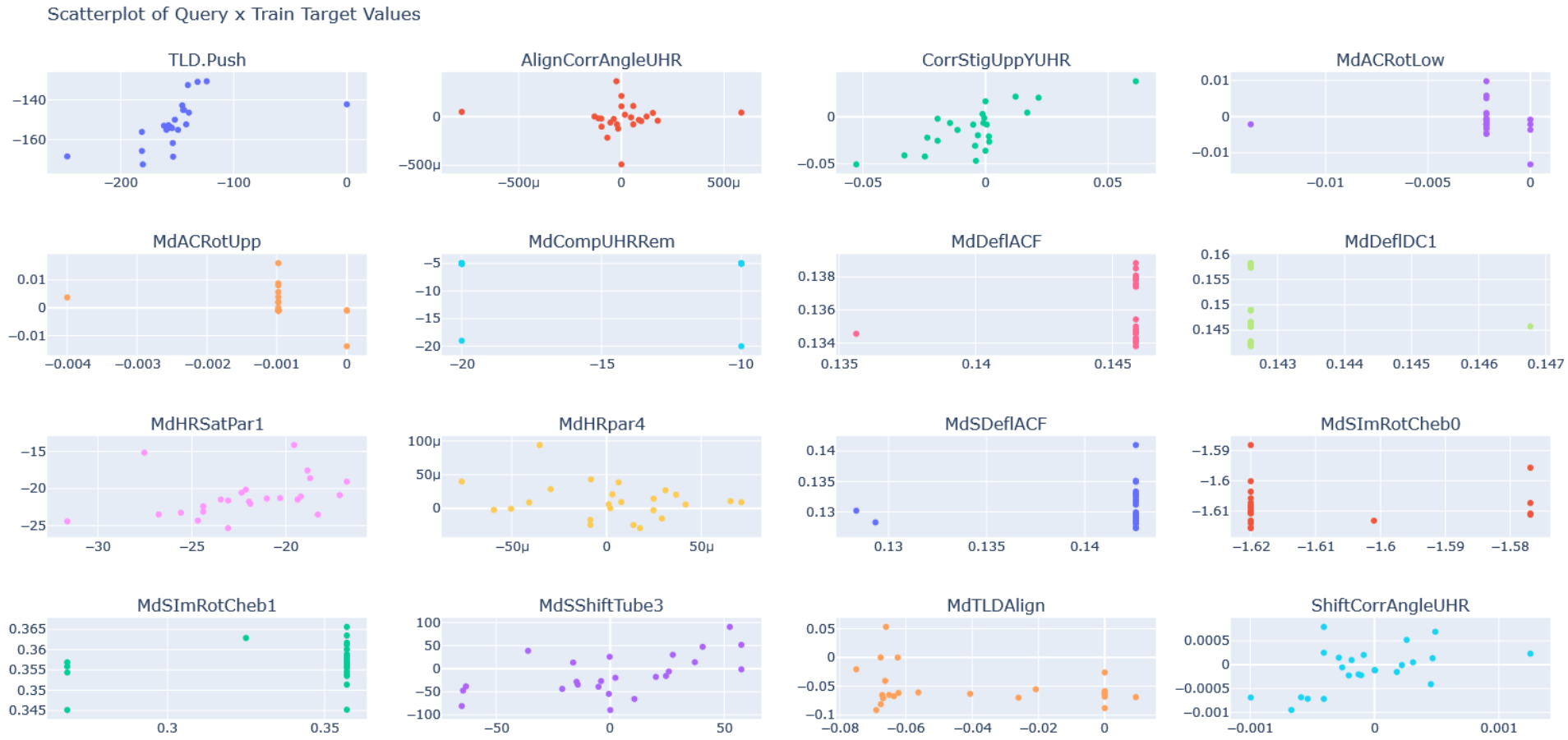
4 – Result and Summary

1 – Task description

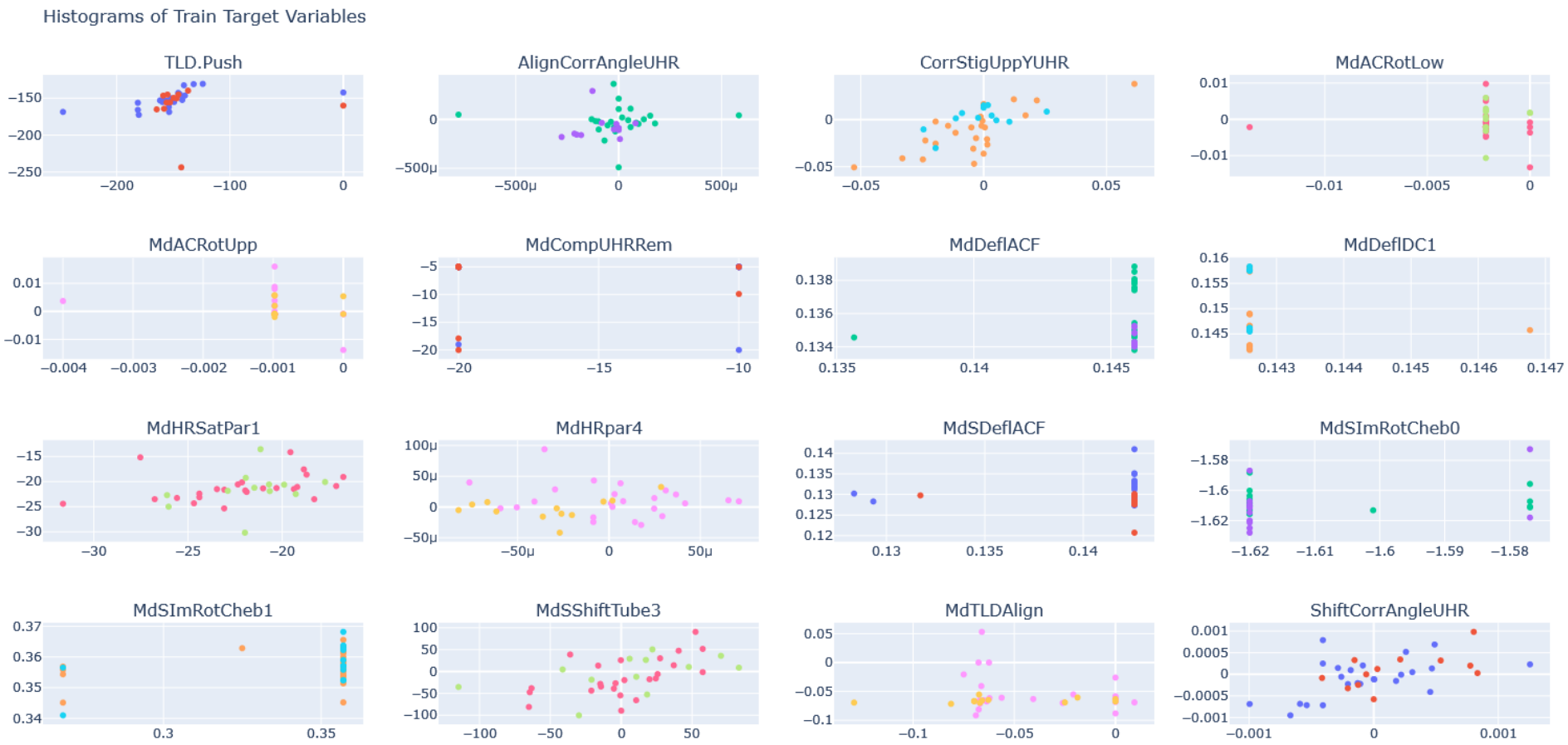
1. Parse input data and perform explorative analysis (data interpretation, visualization, correlations, etc.)
2. Use adequate statistical and/or machine learning methods to match respective target and query column IDs based solely on alignment values (use only the `train` prefixed files)
3. Test your model on the `test` files and perform appropriate evaluation analysis
4. Present the results

2 EDA – Feature Histograms

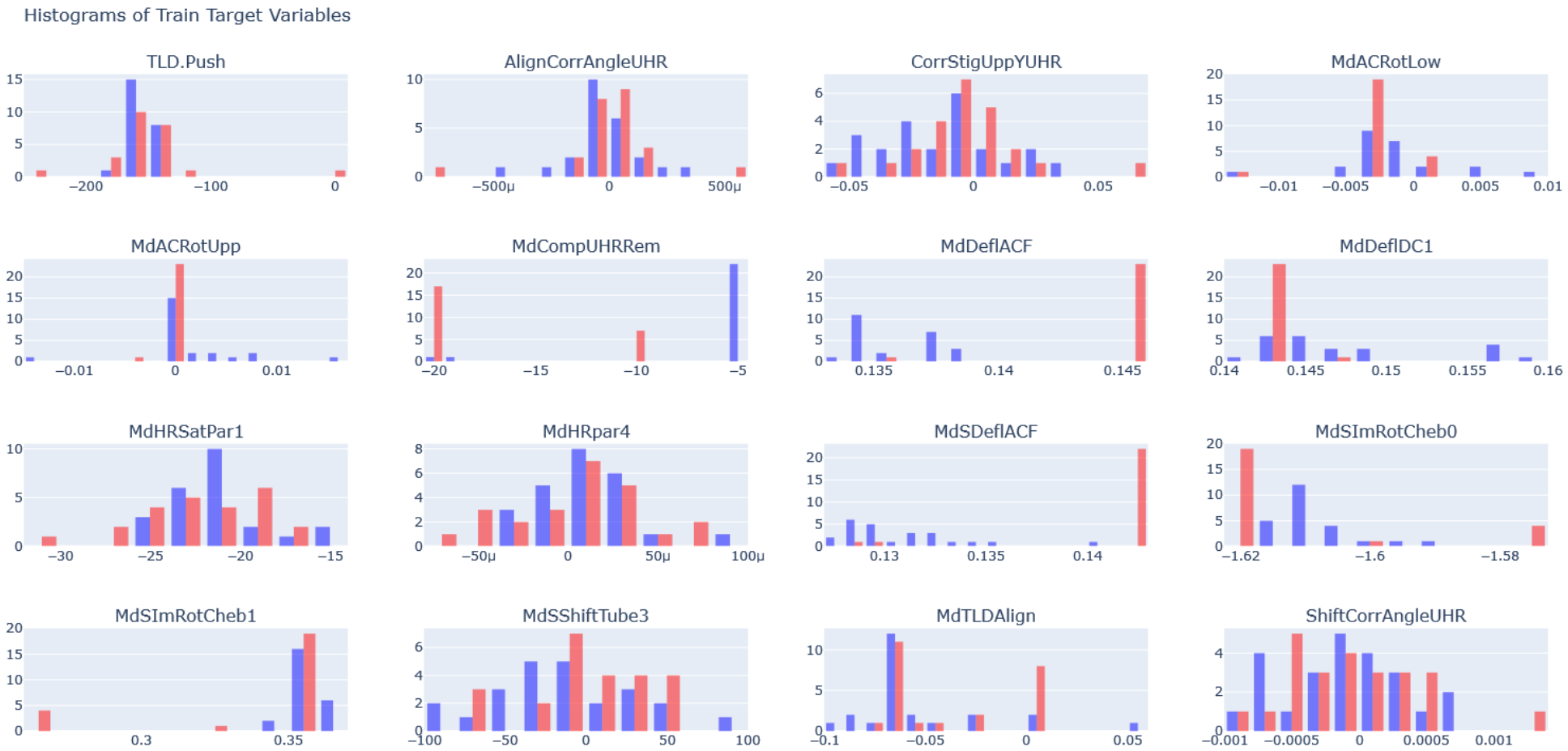
2 EDA – Feature Scatterplots



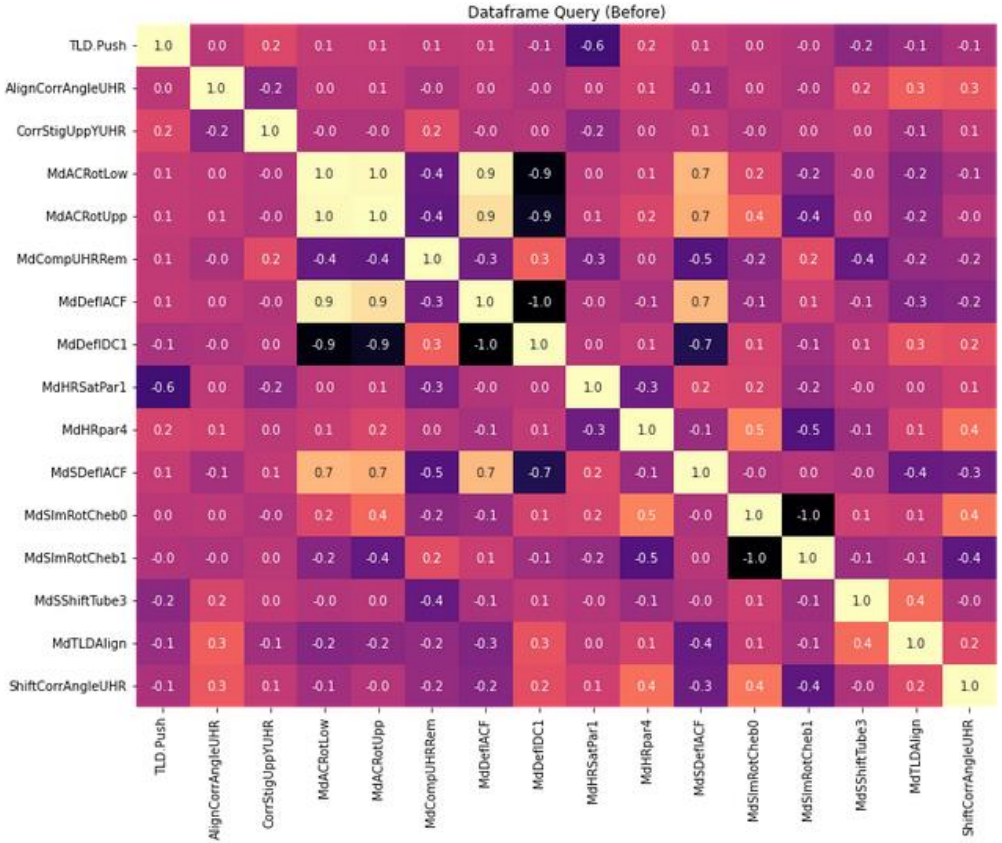
2 EDA – Feature Scatterplots with all data

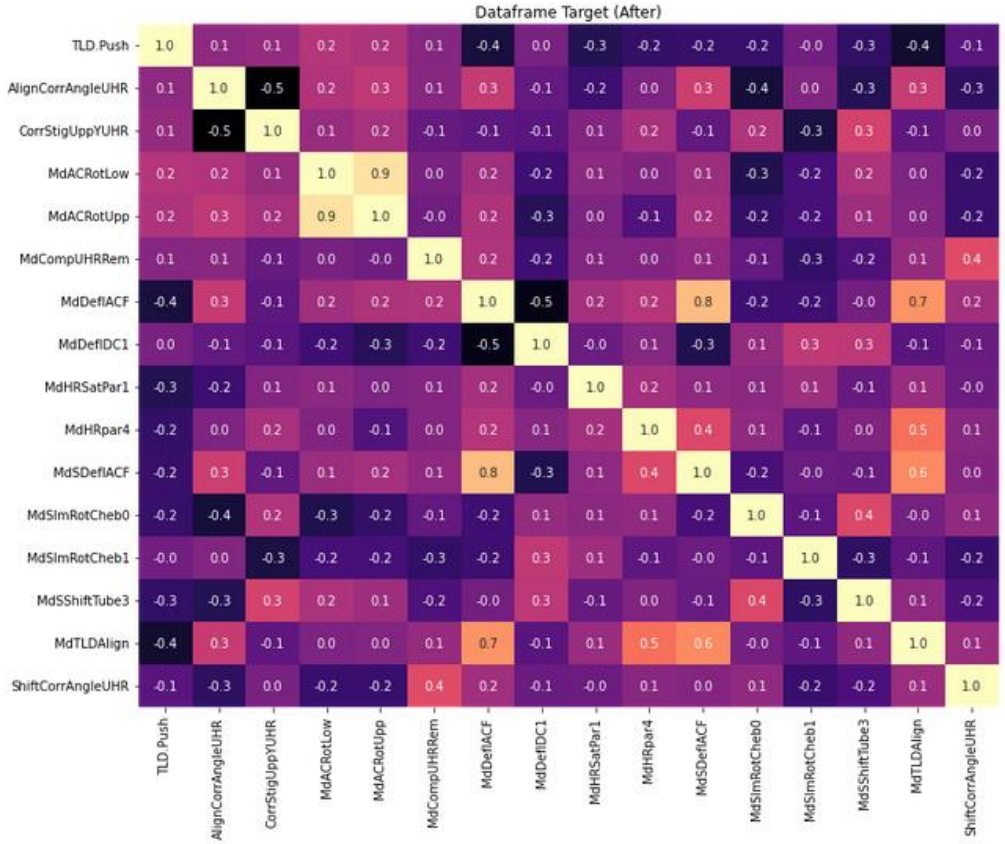


2 EDA – Feature histograms before / after alignment

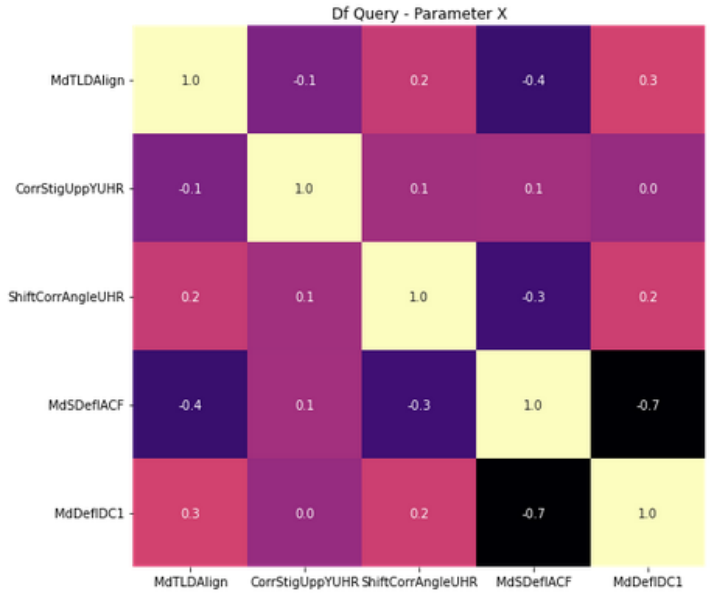


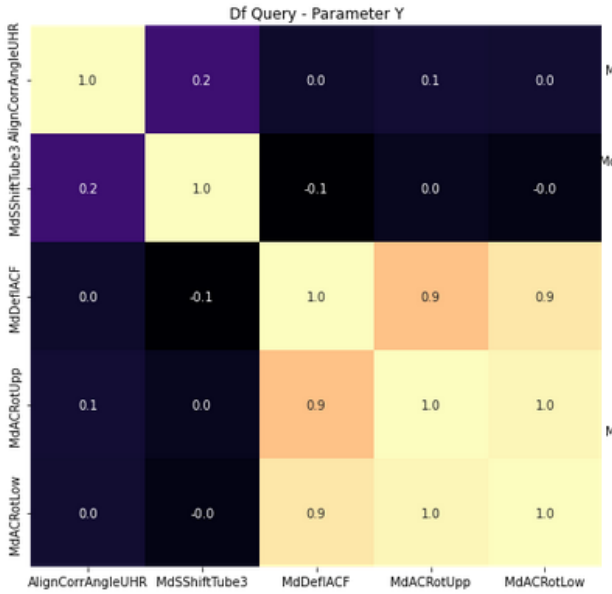
2 EDA – Correlation Heatmaps

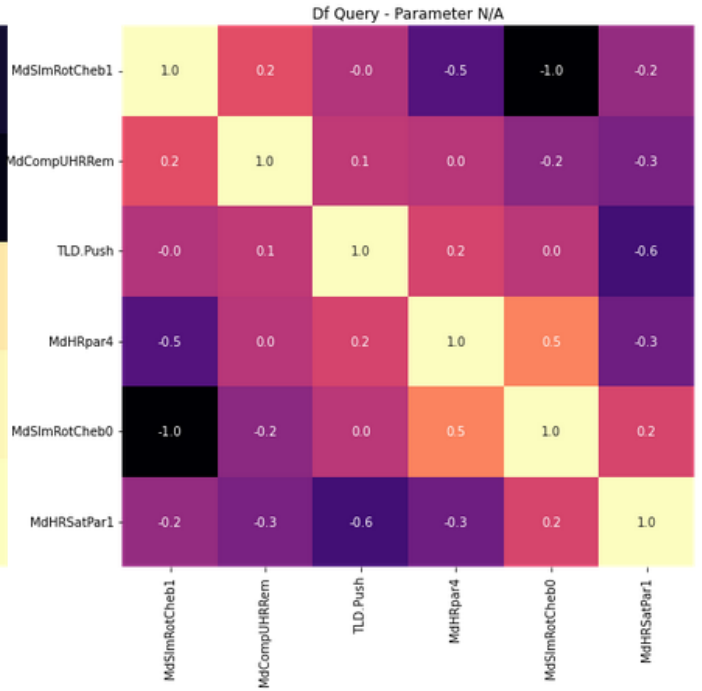




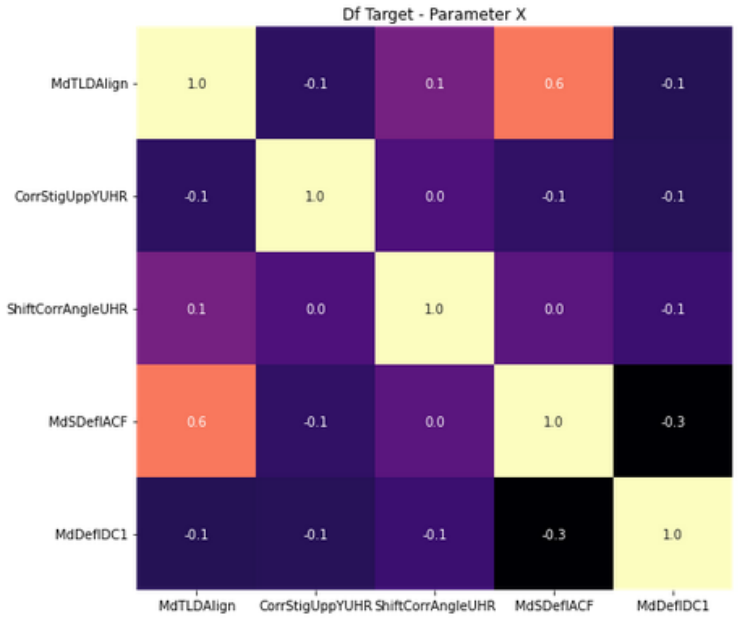
2 EDA – Correlation Heatmaps / Query / Per Parameters

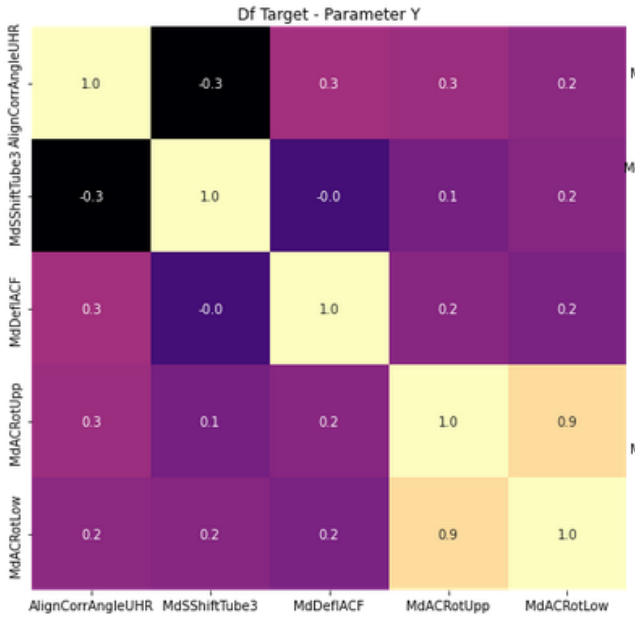


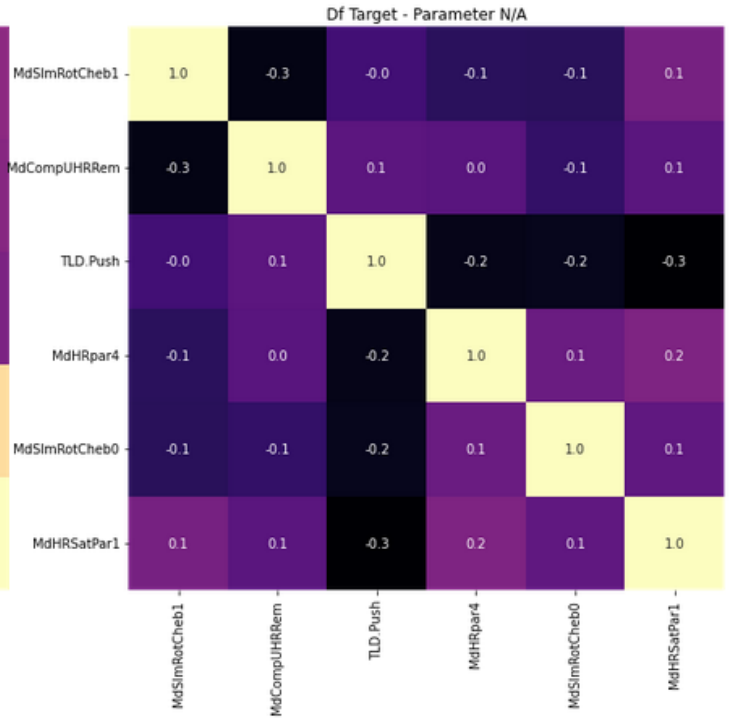




2 EDA – Correlation Heatmaps / Target / Per Parameters

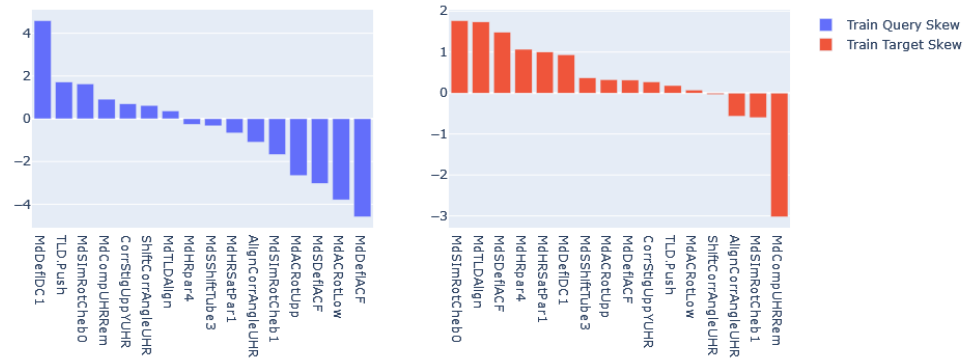




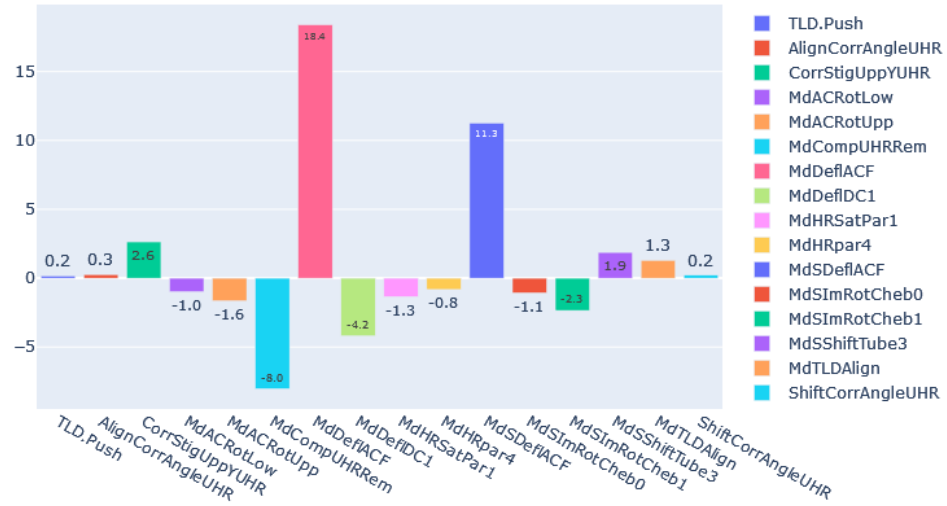


2 EDA – Skew, T-Test, p-values

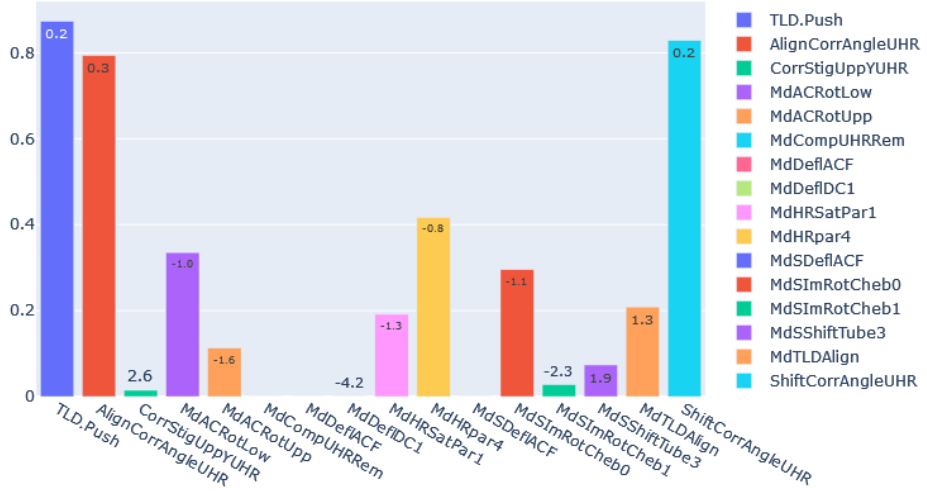
Skew of Train features



Paired t-test results - train dataset



P-values - train dataset



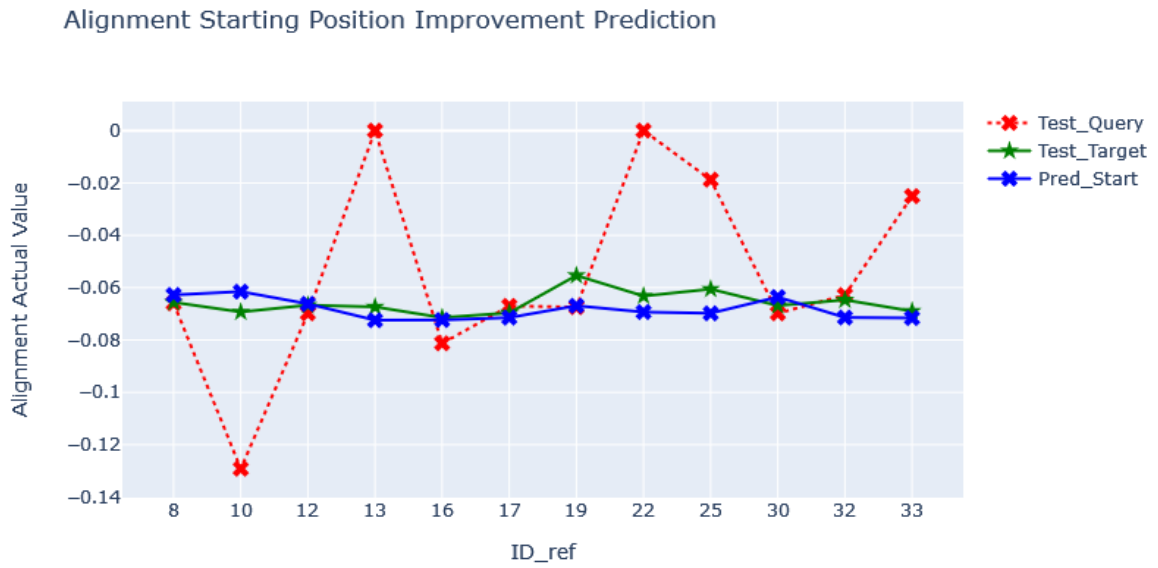
3.1 Parameter **MdTLDAlign**

ML Model : **XGBoost**

```
XGBRegressor(max_depth =3,
       n_estimators = 600,
       learning_rate = 0.1,
       reg_alpha = 0.0001)
```

Feature Selection :

	TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
	X	X	CorrStigUppYUHR	X	X	MdCompUHRRem	X	X	X	X	X	X	X	X	MdTLDAlign	X	MdTLDAlign	XGBoost



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	-0.065936	8	-0.065602	-0.062777	0.000334	0.002825	-745.38	False	Total Manual Alignment w/o ML (measured) :	0.30898
10	-0.129167	10	-0.069280	-0.061492	0.059887	0.007788	87.00	True	Total Manual Alignment with ML (predicted) :	0.05841
12	-0.069792	12	-0.066686	-0.066178	0.003105	0.000508	83.64	True	Total Manual Alignment Difference :	0.25057
13	0.000000	13	-0.067342	-0.072414	0.067342	0.005072	92.47	True	Total Improvement Ratio [%]:	81.1
16	-0.081250	16	-0.071445	-0.072345	0.009805	0.000901	90.81	True	Avg Improvement per piece [%]:	80.79
17	-0.066964	17	-0.069635	-0.071445	0.002671	0.001811	32.20	True		
19	-0.067405	19	-0.055345	-0.066901	0.012060	0.011555	4.19	True		
22	0.000000	22	-0.063194	-0.069319	0.063194	0.006125	90.31	True		
25	-0.018750	25	-0.060548	-0.069767	0.041798	0.009219	77.94	True		
30	-0.069792	30	-0.066850	-0.063575	0.002942	0.003275	-11.33	False		
32	-0.062805	32	-0.064682	-0.071399	0.001877	0.006716	-257.72	False		
33	-0.025000	33	-0.068966	-0.071577	0.043966	0.002612	94.06	True		

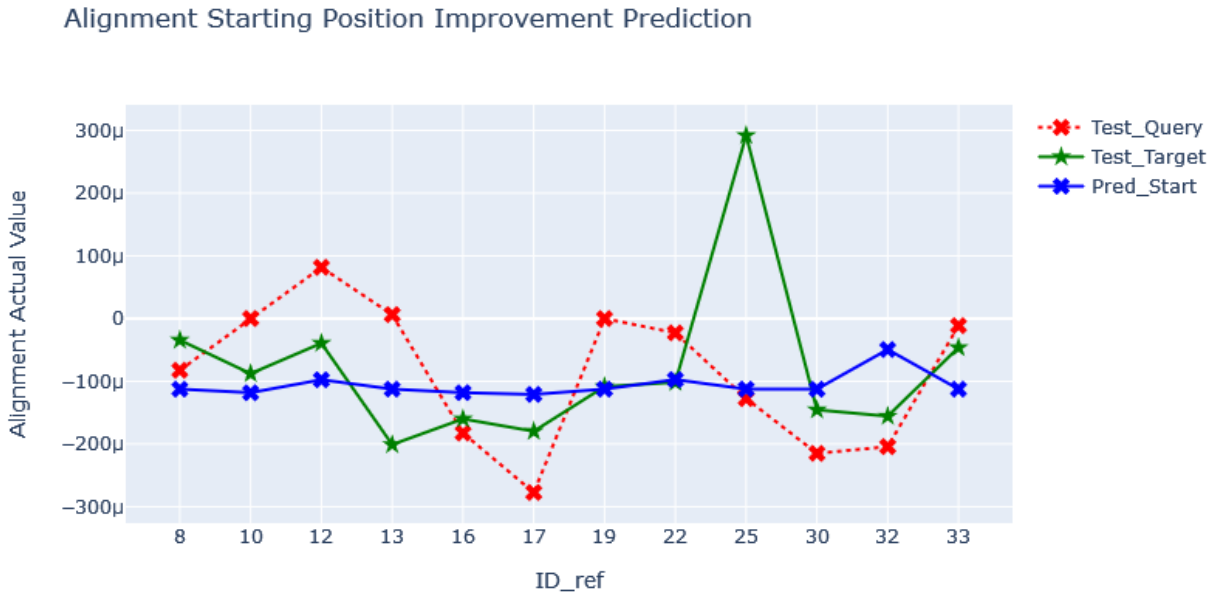
3.2 Parameter **AlignCorrAngleUHR**

ML Model : **RandomForest**

```
RandomForestRegressor(n_estimators = 600,
                      max_depth = 6,
                      min_samples_split = 3,
                      min_samples_leaf = 2,
                      random_state = 42
                      )
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
	X	AlignCorrAngleUHR	CorrStigUppYUHR	X	MdACRotUpp	MdCompUHRRem	X	MdDeflDC1	MdHRSatPar1	MdHRpar4	X	MdSImRotCheb0	MdSImRotCheb1	X	MdTLDAlign	ShiftCorrAngleUHR	AlignCorrAngleUHR	RandomForest



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	-0.000082	8	-0.000034	-0.000112	0.000048	0.000078	-62.57	False	Total Manual Alignment w/o ML (measured) :	0.00134
10	0.000000	10	-0.000088	-0.000118	0.000088	0.000030	65.37	True	Total Manual Alignment with ML (predicted) :	0.00097
12	0.000082	12	-0.000039	-0.000098	0.000121	0.000058	51.83	True	Total Manual Alignment Difference :	0.00037
13	0.000006	13	-0.000201	-0.000112	0.000207	0.000088	57.30	True	Total Improvement Ratio [%]:	27.5
16	-0.000183	16	-0.000160	-0.000118	0.000023	0.000042	-85.43	False	Avg Improvement per piece [%]:	45.82
17	-0.000277	17	-0.000180	-0.000121	0.000098	0.000059	39.81	True		
19	0.000000	19	-0.000108	-0.000112	0.000108	0.000004	96.33	True		
22	-0.000023	22	-0.000102	-0.000097	0.000079	0.000004	94.51	True		
25	-0.000127	25	0.000292	-0.000112	0.000419	0.000404	3.57	True		
30	-0.000215	30	-0.000146	-0.000112	0.000069	0.000033	52.33	True		
32	-0.000204	32	-0.000156	-0.000049	0.000049	0.000106	-119.26	False		
33	-0.000011	33	-0.000046	-0.000112	0.000035	0.000066	-89.65	False		

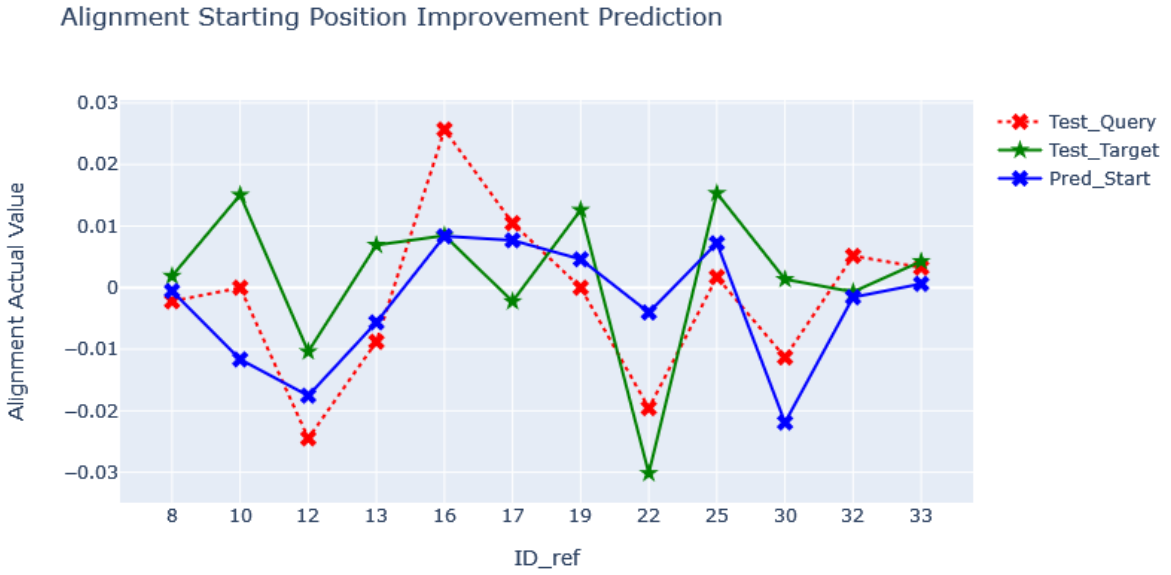
3.3 Parameter **CorrStigUppYUHR**

ML Model : **SupportVectorMachine**

```
= SVR(C= 25,
epsilon= 0.0004,
gamma=0.0004
)
```

Feature Selection :

	TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
TLD.Push		X	CorrStigUppYUHR	X	X	X	X	X	X	MdHRpar4	MdSDeflACF	MdSImRotCheb0	X	MdSShiftTube3	X	ShiftCorrAngleUHR	CorrStigUppYUHR	SVR



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	-0.002219	8	0.001885	-0.000542	0.004104	0.002427	40.86	True	Total Manual Alignment w/o ML (measured) :	0.13529
10	0.000000	10	0.015071	-0.011659	0.015071	0.026730	-77.36	False	Total Manual Alignment with ML (predicted) :	0.12892
12	-0.024475	12	-0.010388	-0.017552	0.014086	0.007164	49.14	True	Total Manual Alignment Difference :	0.00637
13	-0.008784	13	0.006913	-0.005645	0.015697	0.012558	20.00	True	Total Improvement Ratio [%]:	4.71
16	0.025680	16	0.008452	0.008365	0.017228	0.000087	99.49	True	Avg Improvement per piece [%]:	29.24
17	0.010473	17	-0.002242	0.007686	0.012715	0.009928	21.92	True		
19	0.000000	19	0.012620	0.004614	0.012620	0.008006	36.56	True		
22	-0.019562	22	-0.030155	-0.004038	0.010593	0.026117	-146.55	False		
25	0.001726	25	0.015339	0.007232	0.013613	0.008107	40.45	True		
30	-0.011325	30	0.001357	-0.021910	0.012683	0.023267	-83.46	False		
32	0.005170	32	-0.000693	-0.001561	0.005863	0.000868	85.19	True		
33	0.003263	33	0.004278	0.000618	0.001014	0.003659	-260.84	False		

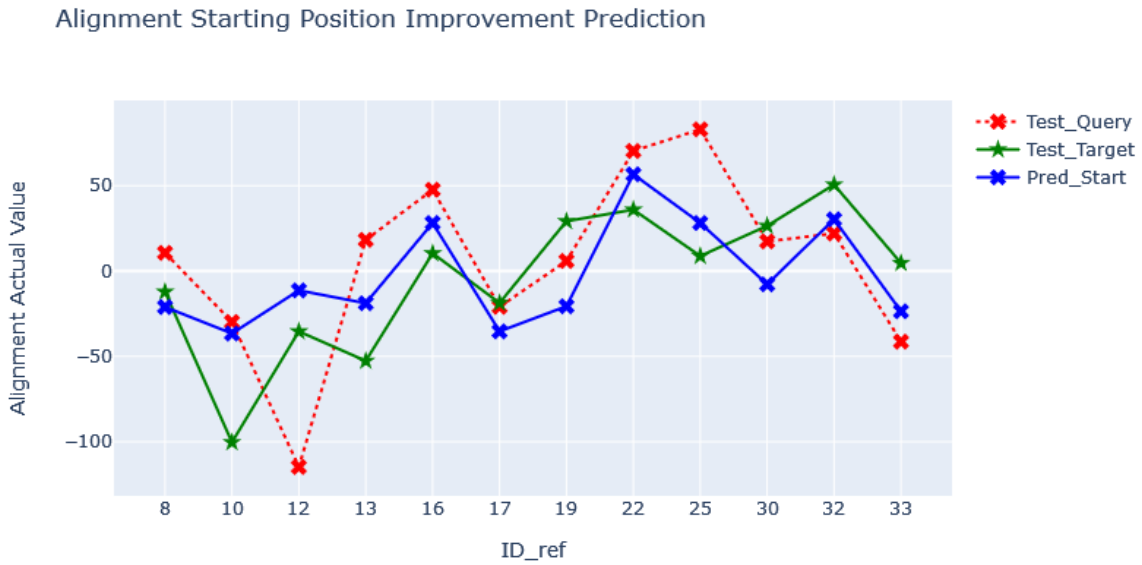
3.4 Parameter **MdSShiftTube3**

ML Model : **XGBoost**

```
XGBRegressor(max_depth =7,
       n_estimators = 600,
       learning_rate = 0.005,
       reg_alpha = 0.00001,
       objective='reg:squarederror'
       )
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
TLD.Push	AlignCorrAngleUHR	X	MdACRotLow	MdACRotUpp	MdCompUHRRem	X	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	X	MdSShiftTube3	XGBoost



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	10.661954	8	-12.157322	-21.069418	22.819276	8.912096	60.94	True	Total Manual Alignment w/o ML (measured) :	498.70344
10	-29.817424	10	-100.327584	-36.579876	70.510160	63.747708	9.59	True	Total Manual Alignment with ML (predicted) :	337.46911
12	-114.771656	12	-35.373122	-11.346097	79.398534	24.027025	69.74	True	Total Manual Alignment Difference :	161.23433
13	18.273001	13	-52.650534	-18.884842	70.923535	33.765692	52.39	True	Total Improvement Ratio [%]:	32.33
16	47.688876	16	10.375920	28.148214	37.312956	17.772294	52.37	True	Avg Improvement per piece [%]:	39.255
17	-20.918074	17	-18.726773	-35.351894	2.191301	16.625121	-658.69	False		
19	5.936010	19	29.190904	-20.629072	23.254894	49.819976	-114.23	False		
22	70.349443	22	35.957977	56.637791	34.391466	20.679814	39.87	True		
25	82.999339	25	8.791515	28.148214	74.207825	19.356700	73.92	True		
30	17.357075	30	26.476125	-7.783036	9.119051	34.259161	-275.69	False		
32	22.041068	32	50.610724	30.333767	28.569656	20.276957	29.03	True		
33	-41.329392	33	4.675398	-23.551168	46.004791	28.226567	38.64	True		

3.5 Parameter **ShiftCorrAngleUHR**

ML Model : **XGBoost**

```
XGBRegressor(max_depth =7,
       n_estimators = 200,
       learning_rate = 0.04,
       reg_alpha = 0.00006,
       reg_lambda = 0.006,
       objective='reg:squarederror'
       )
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	ShiftCorrAngleUHR	ShiftCorrAngleUHR	XGBoost

ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	-0.000126	8	-0.000244	-0.000146	0.000118	0.000098	17.12	True	Total Manual Alignment w/o ML (measured) :	0.00369
10	0.000212	10	0.000345	0.000096	0.000132	0.000248	-88.05	False	Total Manual Alignment with ML (predicted) :	0.00314
12	0.000833	12	0.000028	0.000096	0.000805	0.000068	91.53	True	Total Manual Alignment Difference :	0.00055
13	0.000029	13	0.000123	0.000096	0.000094	0.000027	71.22	True	Total Improvement Ratio [%]:	14.84
16	-0.000208	16	-0.000324	-0.000146	0.000116	0.000179	-53.96	False	Avg Improvement per piece [%]:	-0.94
17	-0.000060	17	0.000000	0.000096	0.000060	0.000096	-59.08	False		
19	0.000000	19	-0.000577	0.000096	0.000577	0.000673	-16.68	False		
22	0.000775	22	0.000200	0.000096	0.000575	0.000104	81.90	True		
25	0.000538	25	0.000321	0.000096	0.000217	0.000225	-3.62	False		
30	0.000803	30	0.000983	0.000096	0.000181	0.000887	-390.70	False		
32	-0.000415	32	-0.000084	-0.000146	0.000331	0.000062	81.32	True		
33	-0.000154	33	0.000329	-0.000146	0.000483	0.000475	1.74	True		

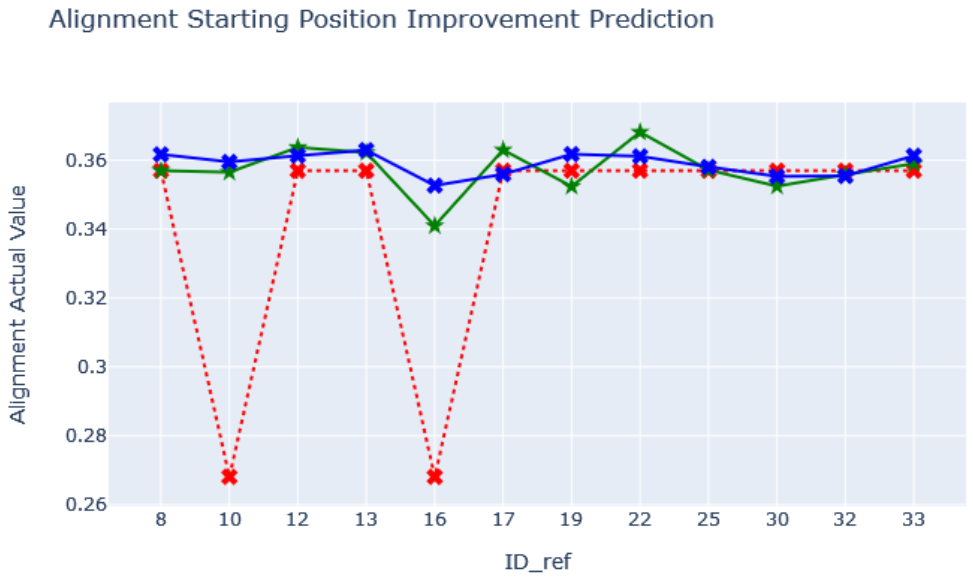
3.6 Parameter **MdSImRotCheb1**

ML Model : **XGBoost**

```
XGBRegressor(max_depth =8,
       n_estimators = 500,
       learning_rate = 0.04,
       reg_alpha = 0.00006,
       reg_lambda = 0.006,
       subsample = 0.9,
       objective='reg:squarederror'
       )
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
TLD.Push	AlignCorrAngleUHR	X	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	X	MdTLDAlign	X	MdSImRotCheb1	XGBoost



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	0.3570	8	0.357000	0.361734	0.000000	0.004734	-inf	False	Total Manual Alignment w/o ML (measured) :	0.20309
10	0.2681	10	0.356543	0.359549	0.088443	0.003005	96.60	True	Total Manual Alignment with ML (predicted) :	0.05244
12	0.3570	12	0.363769	0.361340	0.006769	0.002428	64.12	True	Total Manual Alignment Difference :	0.15065
13	0.3570	13	0.362275	0.362955	0.005275	0.000680	87.10	True	Total Improvement Ratio [%]:	74.18
16	0.2681	16	0.340991	0.352673	0.072891	0.011682	83.97	True	Avg Improvement per piece [%]:	37.02
17	0.3570	17	0.363022	0.355944	0.006022	0.007077	-17.53	False		
19	0.3570	19	0.352428	0.361791	0.004572	0.009362	-104.78	False		
22	0.3570	22	0.368200	0.361184	0.011200	0.007017	37.35	True		
25	0.3570	25	0.357190	0.358119	0.000190	0.000929	-390.17	False		
30	0.3570	30	0.352486	0.355344	0.004514	0.002858	36.69	True		
32	0.3570	32	0.355830	0.355471	0.001170	0.000359	69.28	True		
33	0.3570	33	0.359043	0.361346	0.002043	0.002303	-12.75	False		

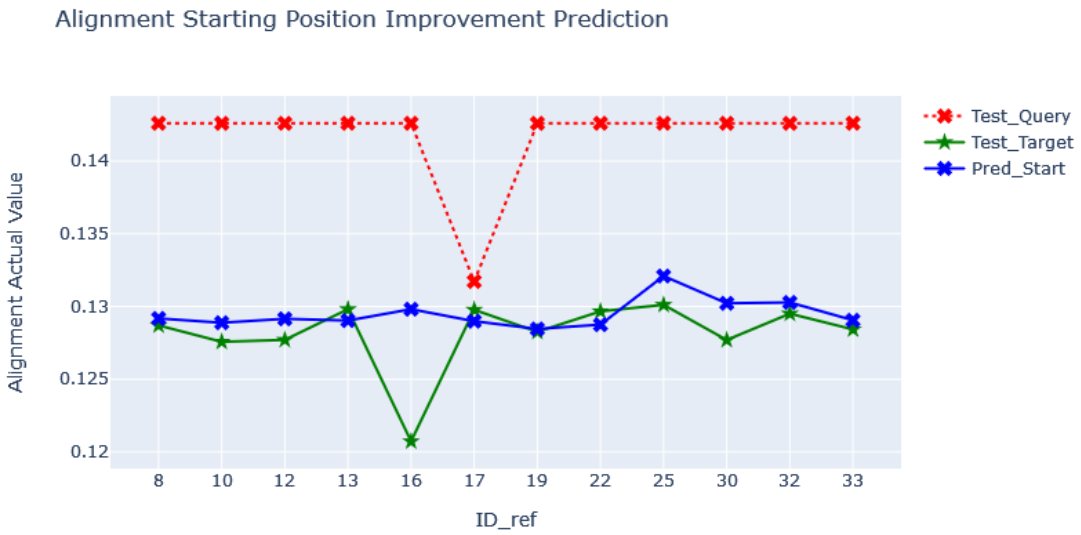
3.7 Parameter **MdSDeflACF**

ML Model : **XGBoost**

```
XGBRegressor(max_depth =8,
       n_estimators = 500,
       learning_rate = 0.04,
       reg_alpha = 0.00006,
       reg_lambda = 0.006,
       objective='reg:squarederror'
       )
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
X	AlignCorrAngleUHR	X	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSlmRotCheb0	MdSlmRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	MdSDeflACF	XGBoost



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	0.142600	8	0.128701	0.129179	0.013899	0.000478	96.56	True	Total Manual Alignment w/o ML (measured) :	0.16236
10	0.142600	10	0.127571	0.128883	0.015029	0.001311	91.28	True	Total Manual Alignment with ML (predicted) :	0.02091
12	0.142600	12	0.127703	0.129154	0.014897	0.001451	90.26	True	Total Manual Alignment Difference :	0.14145
13	0.142600	13	0.129814	0.129035	0.012786	0.000780	93.90	True	Total Improvement Ratio [%]:	87.12
16	0.142600	16	0.120727	0.129814	0.021873	0.009086	58.46	True	Avg Improvement per piece [%]:	92.12
17	0.131723	17	0.129774	0.128989	0.001949	0.000785	59.73	True		
19	0.142600	19	0.128259	0.128458	0.014341	0.000200	98.61	True		
22	0.142600	22	0.129665	0.128754	0.012935	0.000911	92.96	True		
25	0.142600	25	0.130116	0.132095	0.012484	0.001979	84.15	True		
30	0.142600	30	0.127697	0.130223	0.014903	0.002525	83.06	True		
32	0.142600	32	0.129507	0.130279	0.013093	0.000772	94.10	True		
33	0.142600	33	0.128426	0.129056	0.014174	0.000630	95.55	True		

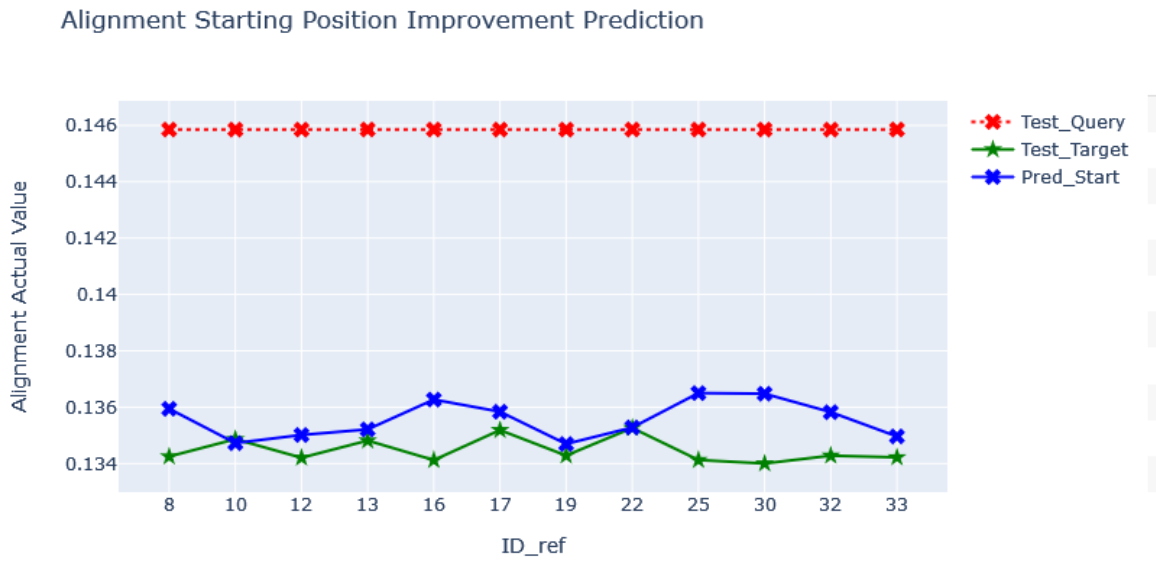
3.8 Parameter **MdDeflACF**

ML Model : **RandomForest**

```
RandomForestRegressor(n_estimators = 600,
                      max_depth = 6,
                      min_samples_split = 3,
                      min_samples_leaf = 2,
                      random_state = 42
                      )
```

Feature Selection :

	TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
	X	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	X	ShiftCorrAngleUHR	MdDeflACF	RandomForest



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	0.145855	8	0.134250	0.135945	0.011605	0.001695	85.39	True	Total Manual Alignment w/o ML (measured) :	0.13664
10	0.145855	10	0.134864	0.134727	0.010991	0.000138	98.75	True	Total Manual Alignment with ML (predicted) :	0.01341
12	0.145855	12	0.134205	0.135014	0.011650	0.000809	93.06	True	Total Manual Alignment Difference :	0.12324
13	0.145855	13	0.134816	0.135212	0.011039	0.000396	96.41	True	Total Improvement Ratio [%]:	90.19
16	0.145855	16	0.134122	0.136269	0.011733	0.002147	81.70	True	Avg Improvement per piece [%]:	93.335
17	0.145855	17	0.135187	0.135840	0.010668	0.000653	93.88	True		
19	0.145855	19	0.134284	0.134695	0.011571	0.000411	96.45	True		
22	0.145855	22	0.135263	0.135279	0.010592	0.000016	99.85	True		
25	0.145855	25	0.134125	0.136501	0.011730	0.002376	79.74	True		
30	0.145855	30	0.134002	0.136479	0.011853	0.002477	79.10	True		
32	0.145855	32	0.134280	0.135825	0.011575	0.001545	86.65	True		
33	0.145855	33	0.134219	0.134962	0.011636	0.000743	93.61	True		

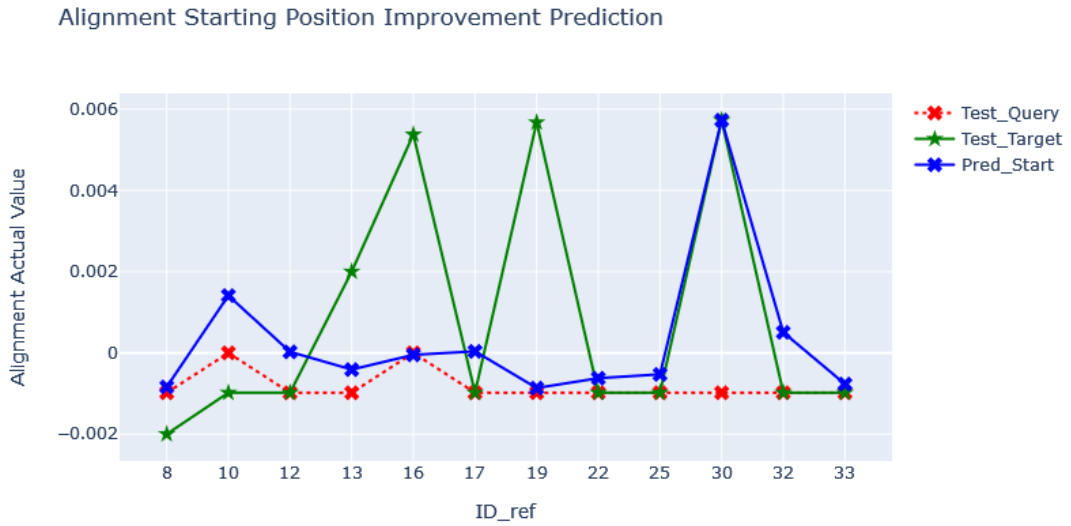
3.9 Parameter **MdACRotUpp**

ML Model : **XGBoost**

```
XGBRegressor(max_depth =7,
       n_estimators = 250,
       learning_rate = 0.08,
       reg_lambda = 0.1000006,
       objective='reg:squarederror',
       subsample = 0.7,
       )
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
X	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	X	X	X	MdDeflDC1	MdHRSatPar1	MdHRpar4	X	X	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	MdACRotUpp	XGBoost



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	-0.00098	8	-0.002000	-0.000839	0.001020	0.001161	-13.81	False	Total Manual Alignment w/o ML (measured) :	0.02372
10	0.00000	10	-0.000980	0.001410	0.000980	0.002390	-143.90	False	Total Manual Alignment with ML (predicted) :	0.02246
12	-0.00098	12	-0.000980	0.000026	0.000000	0.001006	-inf	False	Total Manual Alignment Difference :	0.00126
13	-0.00098	13	0.002000	-0.000411	0.002980	0.002411	19.10	True	Total Improvement Ratio [%]:	5.31
16	0.00000	16	0.005377	-0.000048	0.005377	0.005426	-0.90	False	Avg Improvement per piece [%]:	-inf
17	-0.00098	17	-0.000980	0.000038	0.000000	0.001018	-inf	False		
19	-0.00098	19	0.005675	-0.000865	0.006655	0.006540	1.73	True		
22	-0.00098	22	-0.000980	-0.000623	0.000000	0.000357	-inf	False		
25	-0.00098	25	-0.000980	-0.000529	0.000000	0.000451	-inf	False		
30	-0.00098	30	0.005729	0.005730	0.006709	0.000001	99.98	True		
32	-0.00098	32	-0.000980	0.000506	0.000000	0.001486	-inf	False		
33	-0.00098	33	-0.000980	-0.000767	0.000000	0.000213	-inf	False		

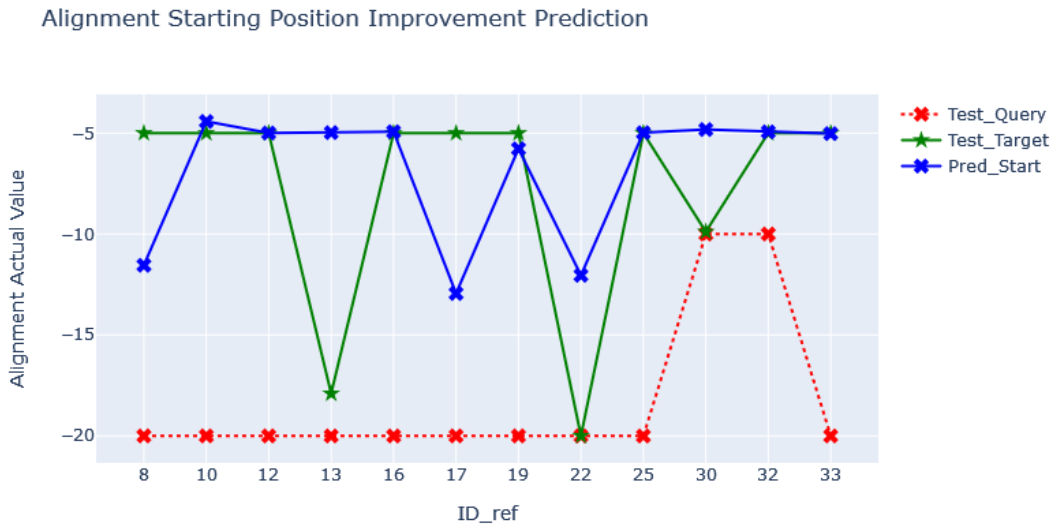
3.10 Parameter **MdCompUHRRem**

ML Model : **XGBoost**

```
XGBRegressor(max_depth =5,
       n_estimators = 250,
       learning_rate = 0.08,
       reg_lambda = 1.5000006,
       objective='reg:squarederror',
       subsample = 0.9
       )
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
TLD.Push	X	CorrStigUppYUHR	X	X	MdCompUHRRem	X	X	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	X	X	X	MdCompUHRRem	XGBoost



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	-20.0	8	-5.000000	-11.547351	15.000000	6.547351	56.35	True	Total Manual Alignment w/o ML (measured) :	127.20913
10	-20.0	10	-5.000000	-4.420431	15.000000	0.579569	96.14	True	Total Manual Alignment with ML (predicted) :	41.97444
12	-20.0	12	-5.000000	-4.994660	15.000000	0.005340	99.96	True	Total Manual Alignment Difference :	85.23469
13	-20.0	13	-17.902400	-4.972751	2.097600	12.929649	-516.40	False	Total Improvement Ratio [%]:	67.0
16	-20.0	16	-5.000000	-4.932362	15.000000	0.067638	99.55	True	Avg Improvement per piece [%]:	95.49
17	-20.0	17	-5.000000	-12.943271	15.000000	7.943271	47.04	True		
19	-20.0	19	-5.000000	-5.774510	15.000000	0.774510	94.84	True		
22	-20.0	22	-20.000000	-12.045472	0.000000	7.954528	-inf	False		
25	-20.0	25	-5.000000	-4.978936	15.000000	0.021064	99.86	True		
30	-10.0	30	-9.888468	-4.820951	0.111532	5.067518	-4443.57	False		
32	-10.0	32	-5.000000	-4.936728	5.000000	0.063272	98.73	True		
33	-20.0	33	-5.000000	-5.020734	15.000000	0.020734	99.86	True		

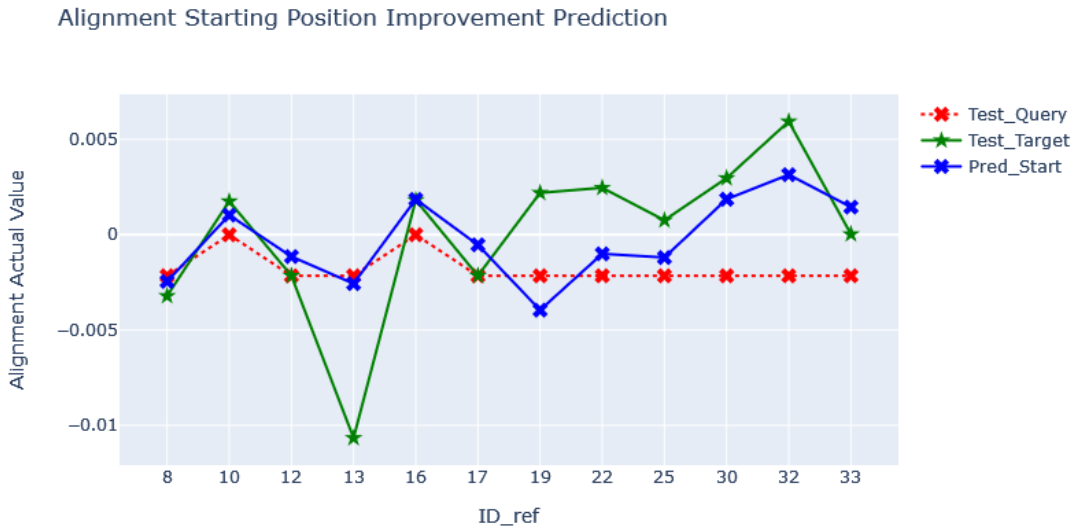
3.11 Parameter **MdACRotLow**

ML Model : **XGBoost**

```
XGBRegressor(max_depth =7,
       n_estimators = 450,
       learning_rate = 0.06,
       reg_lambda = 1.1000006,
       gamma = 0,
       subsample = 0.5,
       colsample_bytree = 0.8,
       objective='reg:squarederror',
       )
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
X	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	X	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	X	ShiftCorrAngleUHR	MdACRotLow	XGBoost



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	-0.00215	8	-0.003211	-0.002451	0.001061	0.000760	28.38	True	Total Manual Alignment w/o ML (measured) :	0.04041
10	0.00000	10	0.001755	0.001026	0.001755	0.000728	58.50	True	Total Manual Alignment with ML (predicted) :	0.02914
12	-0.00215	12	-0.002150	-0.001143	0.000000	0.001007	-inf	False	Total Manual Alignment Difference :	0.01127
13	-0.00215	13	-0.010663	-0.002564	0.008513	0.008099	4.86	True	Total Improvement Ratio [%]:	27.9
16	0.00000	16	0.001817	0.001857	0.001817	0.000041	97.76	True	Avg Improvement per piece [%]:	30.61
17	-0.00215	17	-0.002150	-0.000534	0.000000	0.001616	-inf	False		
19	-0.00215	19	0.002198	-0.003952	0.004348	0.006150	-41.44	False		
22	-0.00215	22	0.002462	-0.000999	0.004612	0.003461	24.96	True		
25	-0.00215	25	0.000761	-0.001194	0.002911	0.001955	32.84	True		
30	-0.00215	30	0.002969	0.001863	0.005119	0.001106	78.40	True		
32	-0.00215	32	0.005942	0.003150	0.008092	0.002792	65.50	True		
33	-0.00215	33	0.000031	0.001454	0.002181	0.001423	34.78	True		

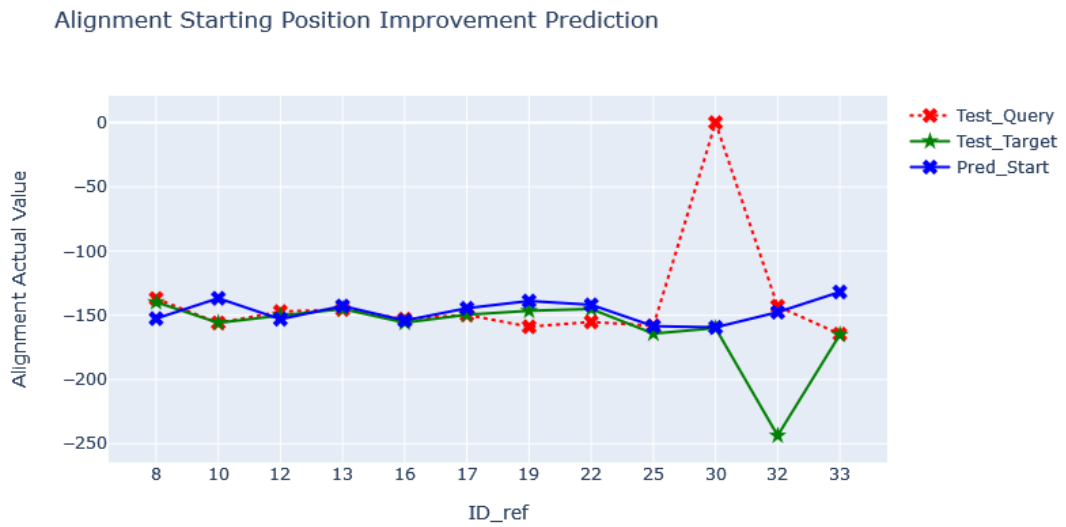
3.12 Parameter **TLD.Push**

ML Model : **XGBoost**

```
XGBRegressor(max_depth =7,
       n_estimators = 450,
       learning_rate = 0.07,
       reg_lambda = 1.0000006,
       gamma = 0.3,
       subsample = 0.6,
       colsample_bytree = 0.75,
       objective='reg:squarederror',
       )
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
X	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	X	X	X	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	TLD.Push	XGBoost



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	-137.000000	8	-139.772539	-152.588867	2.772539	12.816328	-362.26	False	Total Manual Alignment w/o ML (measured) :	298.96689
10	-155.982161	10	-155.876133	-136.730850	0.106027	19.145283	-17956.94	False	Total Manual Alignment with ML (predicted) :	190.36359
12	-147.094097	12	-150.226659	-153.037781	3.132562	2.811122	10.26	True	Total Manual Alignment Difference :	108.6033
13	-145.509374	13	-145.080000	-142.780991	0.429374	2.299009	-435.43	False	Total Improvement Ratio [%]:	36.33
16	-152.963545	16	-155.755148	-154.114914	2.791603	1.640234	41.24	True	Avg Improvement per piece [%]:	4.775
17	-149.680000	17	-149.582690	-144.486450	0.097310	5.096239	-5137.10	False		
19	-158.881825	19	-146.498770	-138.871872	12.383056	7.626898	38.41	True		
22	-155.155639	22	-145.091922	-141.840683	10.063717	3.251239	67.69	True		
25	-158.174192	25	-164.262110	-158.476868	6.087918	5.785242	4.97	True		
30	0.000000	30	-159.987373	-159.418350	159.987373	0.569023	99.64	True		
32	-142.975644	32	-243.650096	-147.587677	100.674452	96.062419	4.58	True		
33	-164.671465	33	-165.112421	-131.851868	0.440956	33.260553	-7442.83	False		

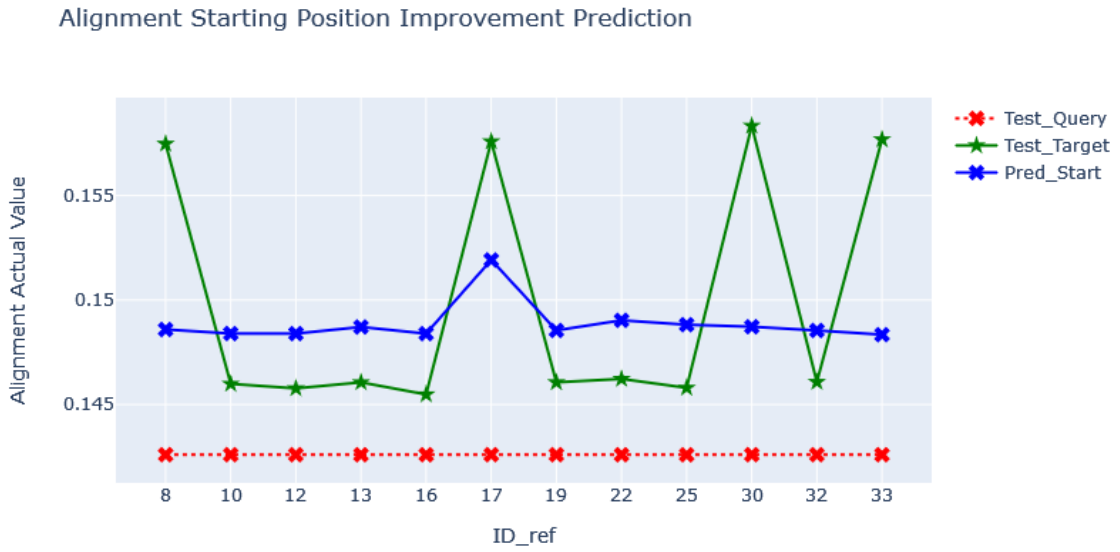
3.13 Parameter **MdDeflDC1**

ML Model : **SVR**

```
= SVR(C= 10,
epsilon= 0.006,
gamma=0.0004
)
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
X	X	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	X	X	X	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	X	MdDeflDC1	SVR



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	0.1426	8	0.157477	0.148592	0.014877	0.008885	40.28	True	Total Manual Alignment w/o ML (measured) :	0.08736
10	0.1426	10	0.145989	0.148397	0.003389	0.002408	28.94	True	Total Manual Alignment with ML (predicted) :	0.05487
12	0.1426	12	0.145773	0.148394	0.003173	0.002621	17.38	True	Total Manual Alignment Difference :	0.03249
13	0.1426	13	0.146061	0.148711	0.003461	0.002650	23.45	True	Total Improvement Ratio [%]:	37.19
16	0.1426	16	0.145483	0.148393	0.002883	0.002911	-0.97	False	Avg Improvement per piece [%]:	28.585
17	0.1426	17	0.157586	0.151919	0.014986	0.005667	62.19	True		
19	0.1426	19	0.146063	0.148548	0.003463	0.002485	28.23	True		
22	0.1426	22	0.146226	0.149030	0.003626	0.002804	22.65	True		
25	0.1426	25	0.145800	0.148821	0.003200	0.003021	5.60	True		
30	0.1426	30	0.158335	0.148724	0.015735	0.009612	38.92	True		
32	0.1426	32	0.146081	0.148546	0.003481	0.002465	29.18	True		
33	0.1426	33	0.157686	0.148344	0.015086	0.009342	38.07	True		

3.14 Parameter **MdHRpar4**

ML Model : **RandomForest**

```
RandomForestRegressor(n_estimators = 600,
                      max_depth = 5,
                      min_samples_split = 2,
                      min_samples_leaf = 2,
                      max_features = 4,
                      random_state = 42
                      )
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
X	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	X	MdHRpar4	RandomForest

ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	-0.000062	8	-0.000007	0.000011	0.000055	1.820658e-05	66.60	True	Total Manual Alignment w/o ML (measured) :	0.00039
10	0.000002	10	0.000010	0.000010	0.000009	4.055519e-07	95.27	True	Total Manual Alignment with ML (predicted) :	0.00018
12	-0.000029	12	-0.000002	0.000006	0.000027	7.877498e-06	70.39	True	Total Manual Alignment Difference :	0.00021
13	-0.000027	13	-0.000042	-0.000001	0.000015	4.050861e-05	-175.29	False	Total Improvement Ratio [%]:	54.31
16	-0.000026	16	-0.000011	0.000010	0.000015	2.041752e-05	-34.20	False	Avg Improvement per piece [%]:	64.875
17	-0.000067	17	0.000008	0.000015	0.000075	7.033840e-06	90.57	True		
19	-0.000075	19	0.000004	0.000009	0.000079	4.907941e-06	93.80	True		
22	-0.000082	22	-0.000005	0.000004	0.000077	9.071066e-06	88.28	True		
25	-0.000036	25	-0.000016	0.000010	0.000021	2.598985e-05	-25.66	False		
30	-0.000020	30	-0.000013	0.000010	0.000007	2.302923e-05	-211.89	False		
32	0.000029	32	0.000033	0.000014	0.000004	1.858326e-05	-349.72	False		
33	-0.000003	33	0.000009	0.000004	0.000012	4.392186e-06	63.15	True		

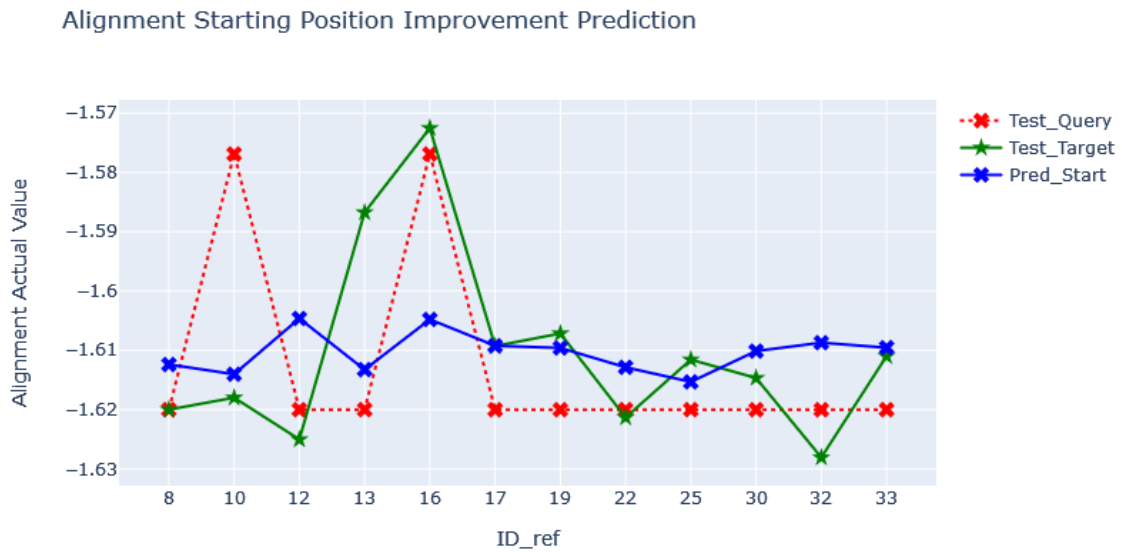
3.15 Parameter **MdSImRotCheb0**

ML Model : **XGBoost**

```
XGBRegressor(max_depth =7,
       n_estimators = 450,
       learning_rate = 0.35,
       reg_lambda = 0.0000006,
       gamma = 0,
       subsample = 0.5,
       colsample_bytree = 0.8,
       objective='reg:squarederror',
       )
```

Feature Selection :

TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
X	X	X	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	MdSImRotCheb0	XGBoost



ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	-1.6200	8	-1.620000	-1.612385	0.000000	0.007615	-inf	False	Total Manual Alignment w/o ML (measured) :	0.1394
10	-1.5769	10	-1.617976	-1.614029	0.041076	0.003947	90.39	True	Total Manual Alignment with ML (predicted) :	0.1307
12	-1.6200	12	-1.625017	-1.604626	0.005017	0.020390	-306.46	False	Total Manual Alignment Difference :	0.0087
13	-1.6200	13	-1.586758	-1.613256	0.033242	0.026498	20.29	True	Total Improvement Ratio [%]:	6.24
16	-1.5769	16	-1.572544	-1.604791	0.004356	0.032247	-640.32	False	Avg Improvement per piece [%]:	17.525
17	-1.6200	17	-1.609282	-1.609231	0.010718	0.000051	99.53	True		
19	-1.6200	19	-1.607162	-1.609598	0.012838	0.002436	81.03	True		
22	-1.6200	22	-1.621317	-1.612862	0.001317	0.008455	-542.04	False		
25	-1.6200	25	-1.611572	-1.615313	0.008428	0.003741	55.61	True		
30	-1.6200	30	-1.614667	-1.610120	0.005333	0.004546	14.76	True		
32	-1.6200	32	-1.628059	-1.608712	0.008059	0.019347	-140.07	False		
33	-1.6200	33	-1.610985	-1.609555	0.009015	0.001430	84.13	True		

3.16 Parameter **MdHRSatPar1**

ML Model : **XGBoost**

```
XGBRegressor(max_depth =8,
       n_estimators = 450,
       learning_rate = 0.08,
       reg_lambda = 0.0000006,
       gamma = 0,
       subsample = 0.5,
       colsample_bytree = 0.8,
       objective='reg:squarederror',
       )
```

Feature Selection :

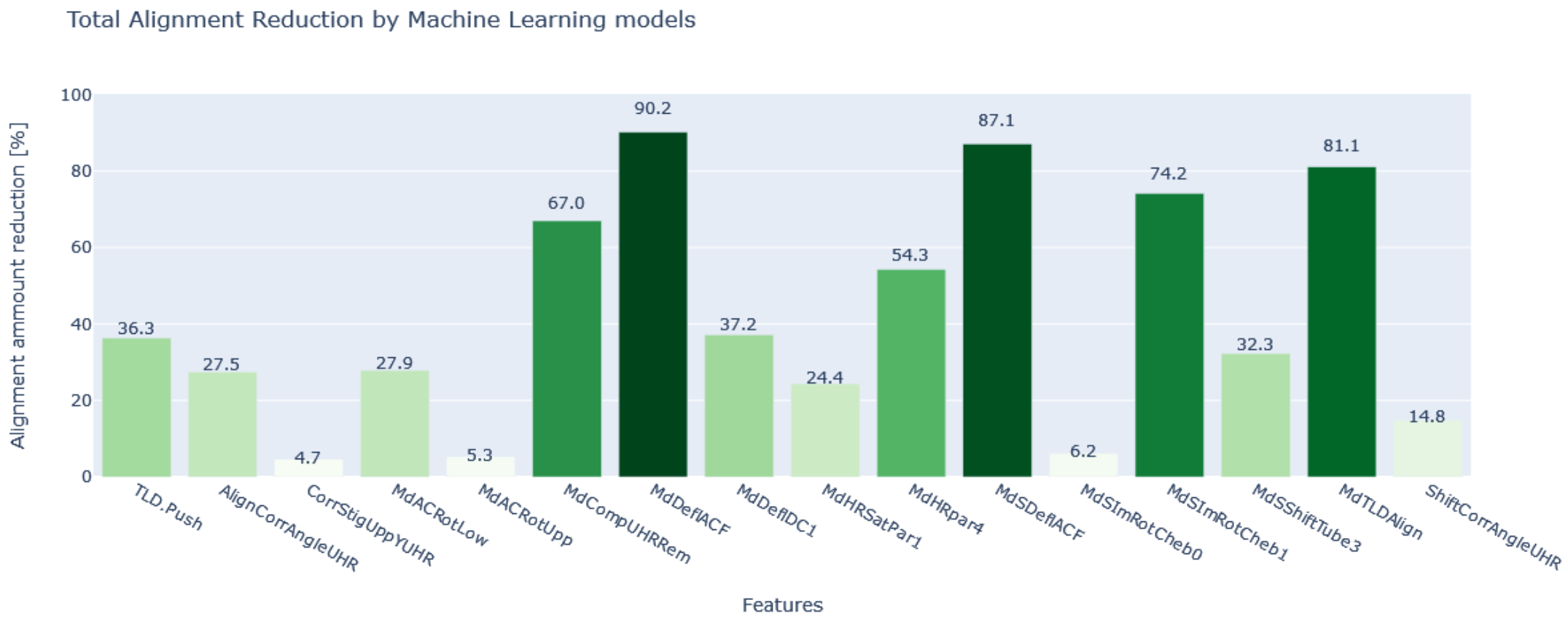
TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
X	AlignCorrAngleUHR	CorrStigUppYUHR	X	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	MdHRSatPar1	XGBoost

ID	Test_Query	ID_ref	Test_Target	Pred_Start	MeasAlgnWrk	PredAlgnWrk	%improved	ML_success	Custom_Metrics	Metrics_Val
8	-19.898864	8	-20.611293	-20.930977	0.712429	0.319684	55.13	True	Total Manual Alignment w/o ML (measured) :	31.83874
10	-22.893730	10	-21.843574	-22.921238	1.050157	1.077664	-2.62	False	Total Manual Alignment with ML (predicted) :	24.06697
12	-21.152948	12	-13.554468	-20.263668	7.598480	6.709200	11.70	True	Total Manual Alignment Difference :	7.77177
13	-21.480564	13	-21.229661	-21.743057	0.250902	0.513396	-104.62	False	Total Improvement Ratio [%]:	24.41
16	-21.936284	16	-19.252956	-22.497530	2.683328	3.244574	-20.92	False	Avg Improvement per piece [%]:	20.585
17	-21.980833	17	-30.164328	-23.862196	8.183495	6.302132	22.99	True		
19	-20.662968	19	-21.887827	-21.436899	1.224859	0.450928	63.19	True		
22	-19.282921	22	-22.470475	-19.862259	3.187554	2.608216	18.18	True		
25	-17.723630	25	-20.074173	-21.083210	2.350542	1.009037	57.07	True		
30	-26.115909	30	-22.692141	-22.679779	3.423768	0.012362	99.64	True		
32	-26.031764	32	-24.989201	-23.174688	1.042562	1.814513	-74.04	False		
33	-20.702045	33	-20.571385	-20.576649	0.130660	0.005263	95.97	True		

4 - Final Feature Selection Overview

	TLD.Push	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	Target Feature	Model
0	X	X	CorrStigUppYUHR	X	X	MdCompUHRRem	X	X	X	X	X	X	X	X	MdTLDAlign	X	MdTLDAlign	XGBoost
1	X	AlignCorrAngleUHR	CorrStigUppYUHR	X	MdACRotUpp	MdCompUHRRem	X	MdDeflDC1	MdHRSatPar1	MdHRpar4	X	MdSImRotCheb0	MdSImRotCheb1	X	MdTLDAlign	ShiftCorrAngleUHR	AlignCorrAngleUHR	RandomForest
2	TLD.Push	X	CorrStigUppYUHR	X	X	X	X	X	X	MdHRpar4	MdSDeflACF	MdSImRotCheb0	X	MdSShiftTube3	X	ShiftCorrAngleUHR	CorrStigUppYUHR	SVR
3	TLD.Push	AlignCorrAngleUHR	X	MdACRotLow	MdACRotUpp	MdCompUHRRem	X	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	X	MdSShiftTube3	XGBoost
4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	ShiftCorrAngleUHR	ShiftCorrAngleUHR	XGBoost
5	TLD.Push	AlignCorrAngleUHR	X	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	X	MdTLDAlign	X	MdSImRotCheb1	XGBoost
6	X	AlignCorrAngleUHR	X	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	MdSDeflACF	XGBoost
7	X	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	X	ShiftCorrAngleUHR	MdDeflACF	RandomForest
8	X	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	X	X	X	MdDeflDC1	MdHRSatPar1	MdHRpar4	X	X	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	MdACRotUpp	XGBoost
9	TLD.Push	X	CorrStigUppYUHR	X	X	MdCompUHRRem	X	X	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	X	X	X	MdCompUHRRem	XGBoost
10	X	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	X	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	X	ShiftCorrAngleUHR	MdACRotLow	XGBoost
11	X	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	X	X	X	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	TLD.Push	XGBoost
12	X	X	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	X	X	X	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	X	MdDeflDC1	SVR
13	X	AlignCorrAngleUHR	CorrStigUppYUHR	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	X	MdHRpar4	RandomForest
14	X	X	X	MdACRotLow	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	MdSImRotCheb0	XGBoost
15	X	AlignCorrAngleUHR	CorrStigUppYUHR	X	MdACRotUpp	MdCompUHRRem	MdDeflACF	MdDeflDC1	MdHRSatPar1	MdHRpar4	MdSDeflACF	MdSImRotCheb0	MdSImRotCheb1	MdSShiftTube3	MdTLDAlign	ShiftCorrAngleUHR	MdHRSatPar1	XGBoost

4 - Alignment Reduction Results



4 - Future improvement points

- Gain domain experience for better understanding of Feature Engineering
- Get more data from more processes for the ML algorithms
- With more data, apply Neural Networks
- Explore more Machine learning and Pre-processing methods
- Perform more sophisticated hyperparameter tuning with bigger computing power

Thank you for attention