VistaMilk

A World Leading SFI Research Centre

Science Foundation Ireland SFI For what's next

MLGIG Team - Time Series Approaches

Thach Le Nguyen

Postdoctoral Researcher

VistaMilk International Workshop on Spectroscopy and Chemometrics 2021

HOST INSTITUTION

Teagasc Agriculture and Food Development Authority

PARTNER INSTITUTIONS

DCU, UCD Dublin, NUI Galway OE Gaillimh, Waterford Institute of Technology, Tyndall National Institute, ICBF.com, UCC University College Cork Ireland, TSSG, Insight

FUNDED BY:

An Roinn Talmhaíochta, Bia agus Mara Department of Agriculture, Food and the Marine; Ireland's European Structural and Investment Funds Programmes 2014-2020 Co-funded by the Irish Government and the European Union; European Union European Regional Development Fund; Science Foundation Ireland SFI For what's next

About me

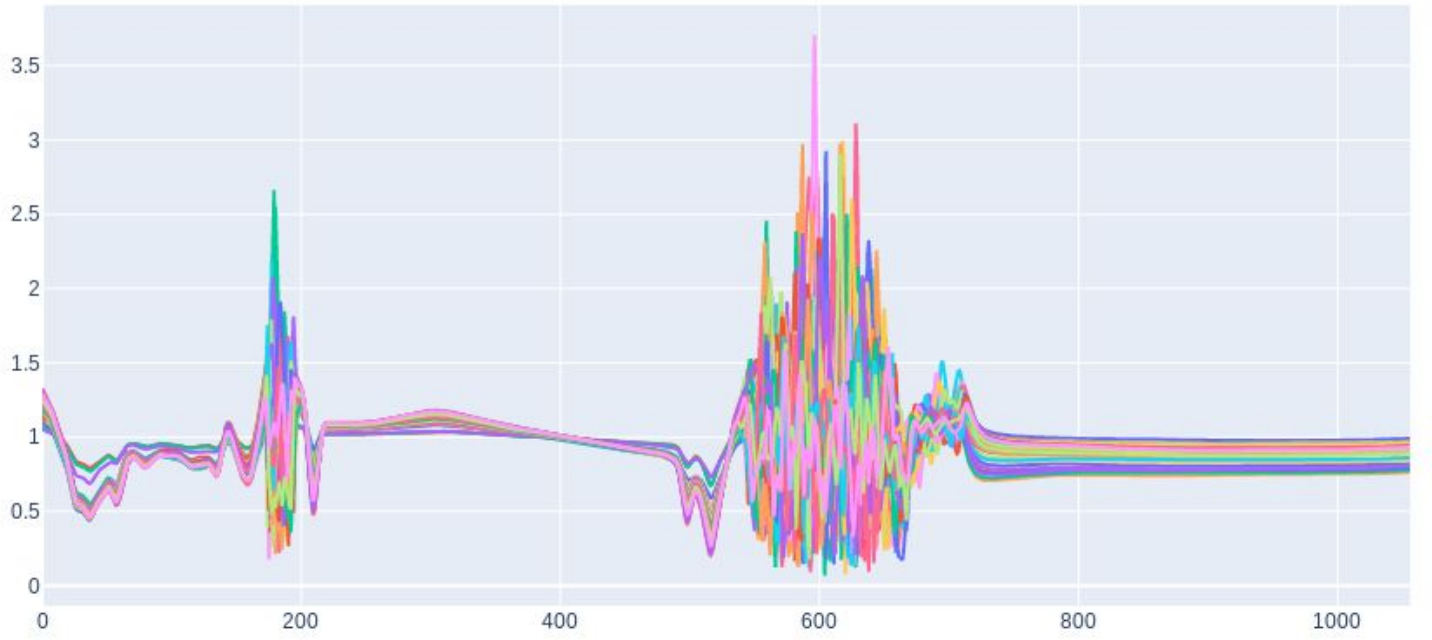
- 2019 - Now : Postdoc Researcher with VistaMilk
- 2015 - 2019: PhD, Computer Science, University College Dublin, Ireland
- 2012 - 2014: MSc, Media Informatic, University of Trento (Italy), RWTH Aachen University (Germany)
- 2005-2010: BSc, Computer Science, HCM University of Technology, Vietnam

Time Series Approaches

Take into account the sequential order of waves

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ippa_case	n_micelle	Native_pH	wave_1	wave_2	wave_3	wave_4	wave_5	wave_6	wave_7	wave_8	wave_9	wave_10	wave_11	wave_12	wave_13
2	1.357356	159.6	7.02	1.066316	1.063704	1.060935	1.057821	1.054414	1.050886	1.04746	1.044318	1.041482	1.038738	1.035657	1.031744	1.026658
3	1.642316	170.7	7.06	1.067492	1.064005	1.060931	1.057989	1.055052	1.052043	1.048972	1.04589	1.042774	1.039398	1.035312	1.029973	1.02301
4	1.93715	163	6.76	1.120782	1.114801	1.10903	1.103285	1.097662	1.092269	1.087206	1.08247	1.077826	1.072701	1.06624	1.057563	1.046167
5	2.552068	143.4	6.57	1.212182	1.200823	1.190243	1.180253	1.170991	1.162457	1.154522	1.146833	1.138638	1.128674	1.115312	1.097051	1.073213
6	2.713626	186.9	6.65	1.237251	1.219401	1.204874	1.192579	1.181946	1.172529	1.164142	1.156589	1.149215	1.140625	1.128857	1.112026	1.08917
7	2.748992	198.3		1.345429	1.330373	1.315704	1.300587	1.284994	1.269116	1.253306	1.23785	1.222615	1.206786	1.188923	1.16746	1.141476
8	2.87265	138.6	6.93	1.109746	1.10617	1.101996	1.097182	1.091973	1.086692	1.081642	1.07696	1.072455	1.067522	1.061231	1.052634	1.041191
9	2.918938	182.1	6.74	1.187982	1.177251	1.167172	1.157682	1.148946	1.141001	1.133757	1.126893	1.119686	1.110902	1.098961	1.08241	1.060593
10	3.178458	188.4	6.65	1.204409	1.193434	1.183228	1.173487	1.164285	1.155647	1.147539	1.139741	1.131625	1.122021	1.109349	1.092119	1.069621
11	3.40232	164.7	6.567	1.255412	1.233578	1.214185	1.19776	1.184288	1.173023	1.16304	1.153481	1.143477	1.131947	1.117578	1.099131	1.075978
12	3.429446	221.3	6.59	1.220451	1.212043	1.202947	1.193197	1.182997	1.172592	1.162236	1.152028	1.141648	1.130161	1.116108	1.097969	1.074875
13	3.454	145.6	6.51	1.245644	1.232289	1.220355	1.209064	1.198217	1.187734	1.177639	1.16786	1.157908	1.146653	1.132441	1.113617	1.089309
14	3.482758	180.4	6.54	1.222205	1.210439	1.20018	1.191115	1.182767	1.174564	1.166059	1.156922	1.146719	1.134668	1.119612	1.100379	1.076395
15	3.494252	178.3	6.68	1.213193	1.202964	1.193934	1.185481	1.17725	1.168958	1.160482	1.151722	1.14232	1.131416	1.117686	1.099768	1.076946
16	3.530988	186.4	6.64	1.168811	1.162572	1.155263	1.147748	1.140999	1.135638	1.131734	1.128636	1.12494	1.118699	1.107904	1.091137	1.068164
17	3.578534	176.5	6.72	1.209157	1.199899	1.191079	1.182105	1.172951	1.163759	1.154725	1.145866	1.136731	1.126248	1.11288	1.095148	1.072347
18	3.589	186.6	6.58	1.246081	1.232543	1.220573	1.209569	1.199226	1.189338	1.179762	1.17026	1.160253	1.148664	1.134043	1.115023	1.090981
19	3.590028	216	6.767	1.161941	1.152823	1.145099	1.137961	1.131025	1.123946	1.11652	1.108604	1.099886	1.089667	1.076876	1.060413	1.039695
20	3.59648	177	6.6	1.246279	1.233477	1.221329	1.209434	1.197954	1.187048	1.176796	1.167019	1.157057	1.145653	1.131145	1.111998	1.087532
21	3.621304	177.2	6.72	1.17089	1.162751	1.154638	1.146487	1.138574	1.131119	1.124203	1.117614	1.110668	1.102138	1.090447	1.074167	1.052676
22	3.621884	178.4	6.965	1.159137	1.150156	1.142534	1.135956	1.130306	1.125169	1.119977	1.114155	1.107185	1.098579	1.087873	1.074741	1.05921
23	3.681504	176	6.88	1.123048	1.115868	1.109475	1.103578	1.098127	1.093078	1.088397	1.083975	1.079467	1.074179	1.067132	1.05736	1.044349
24	3.683046	189.7	6.62	1.164501	1.155739	1.14729	1.138752	1.129828	1.120963	1.112665	1.105218	1.098244	1.090506	1.08015	1.06537	1.04522
25	3.703966	273.9	6.683	1.211849	1.20079	1.188078	1.174473	1.161042	1.148634	1.137741	1.128296	1.119489	1.109785	1.097271	1.0803	1.058222
26	3.710616	110.7	6.65	1.237532	1.225069	1.214012	1.203726	1.193845	1.184187	1.174718	1.165339	1.155576	1.144396	1.13033	1.11196	1.088611
27	3.720848		6.85	1.143415	1.136109	1.12887	1.121852	1.115431	1.109825	1.105004	1.100563	1.09558	1.088592	1.077848	1.061847	1.040018
28	3.730174	146.9	6.73	1.149311	1.1407	1.13236	1.124423	1.11719	1.110796	1.105185	1.100008	1.094462	1.087234	1.076684	1.061343	1.040579
29	3.750444	191.8	6.557	1.250526	1.230053	1.212659	1.198176	1.186417	1.176738	1.168299	1.160148	1.151159	1.139992	1.125245	1.105847	1.081557
30	3.763068	209.1	6.49	1.262811	1.248934	1.236211	1.224089	1.212352	1.200972	1.190052	1.1796	1.169202	1.157815	1.143876	1.125796	1.102701
31	3.763918	170.8	6.64	1.281971	1.26962	1.257737	1.245605	1.233191	1.220645	1.208234	1.196117	1.184031	1.171062	1.155736	1.136501	1.112472
32	3.787046	219.9	6.76	1.184759	1.174984	1.165958	1.157623	1.150018	1.143029	1.136465	1.129988	1.122922	1.114113	1.102056	1.085358	1.06341
33	3.792126	145.5	6.58	1.298394	1.285477	1.273061	1.260329	1.247181	1.233752	1.220345	1.207189	1.194076	1.180105	1.163761	1.143426	1.118171
34	3.805638	160.7	6.73	1.243221	1.232102	1.221403	1.210729	1.200149	1.189775	1.179728	1.169969	1.160038	1.148895	1.135056	1.117086	1.094294
35	3.829784	238.5	6.65	1.218872	1.208716	1.198907	1.188884	1.178609	1.168215	1.157963	1.148028	1.138194	1.127647	1.115069	1.099087	1.078942
36	3.838062	169.6	6.57	1.228325	1.216669	1.205898	1.195556	1.185598	1.175981	1.166688	1.15758	1.148136	1.137262	1.123407	1.10505	1.08143

Train set | Test set

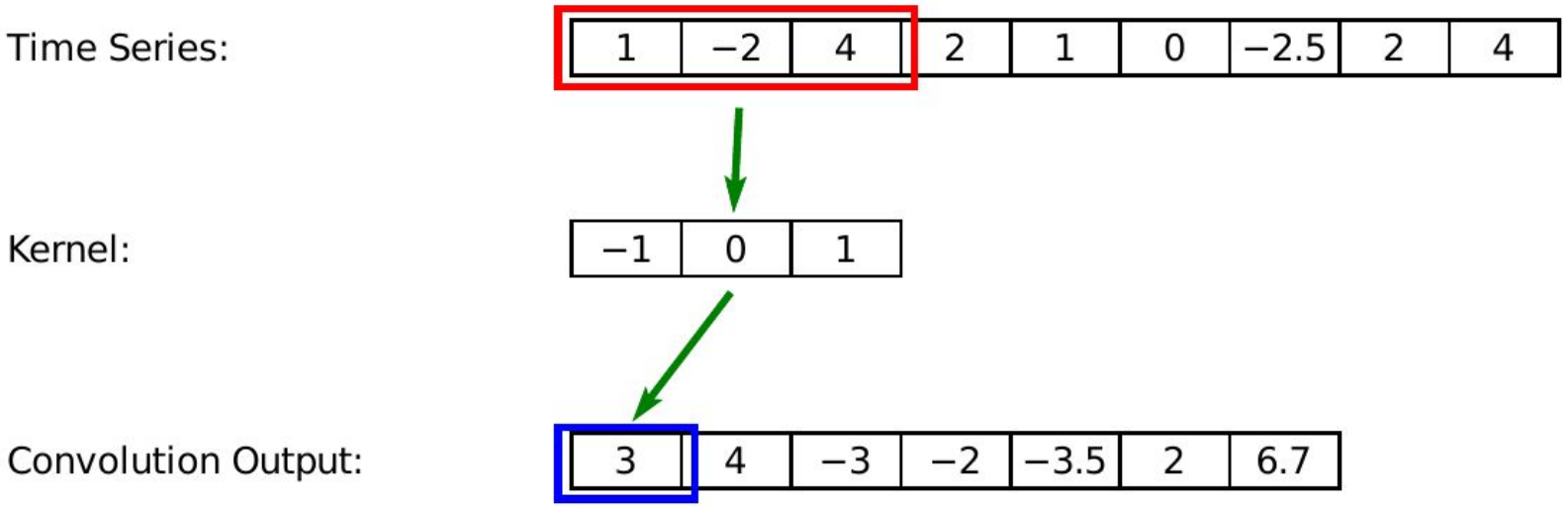


This approach has been successful in the classification domain.
(http://www.timeseriesclassification.com/dataset.php)

ROCKET Time Series Classifier/ Regressor

Angus Dempster, Francois Petitjean, and Geoffrey I Webb - ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels - Data Mining and Knowledge Discovery / arXiv:1910.13051

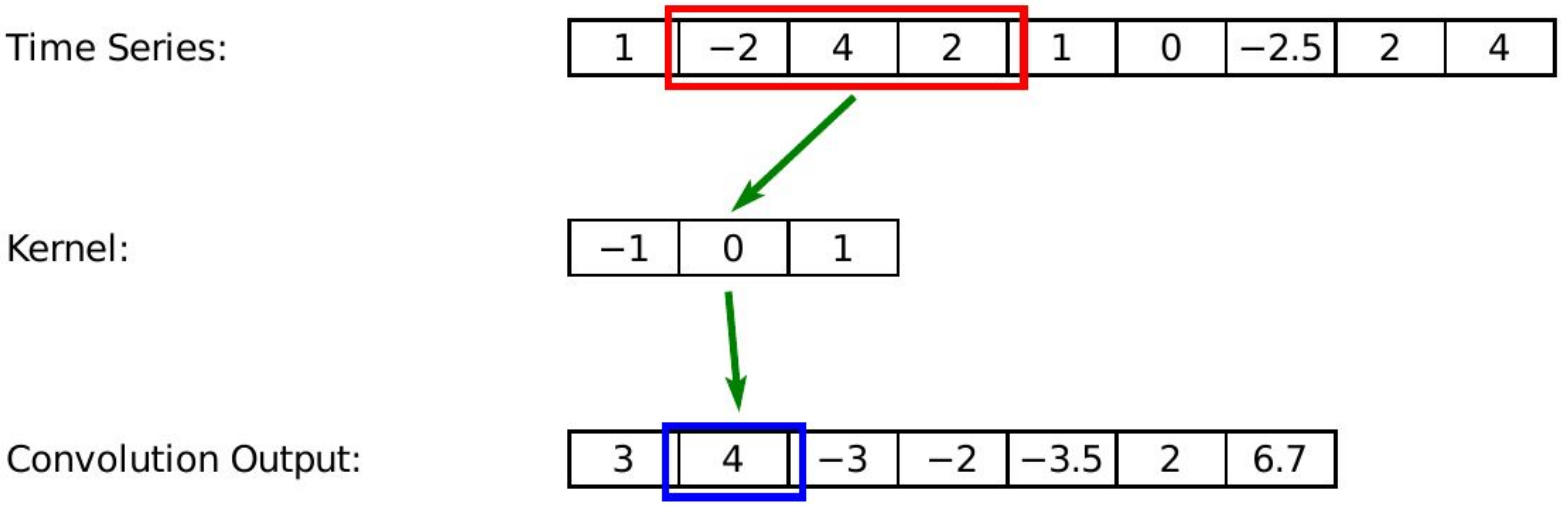
- State-of-the-art time series classifier in terms of both speed and accuracy.
- Applicable to regression problems.
- Key idea: Extract features from time series using random convolution kernels + off-the-shelf ML (RidgeCV)



ROCKET Time Series Classifier/ Regressor

Angus Dempster, Francois Petitjean, and Geoffrey I Webb - ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels - Data Mining and Knowledge Discovery / arXiv:1910.13051

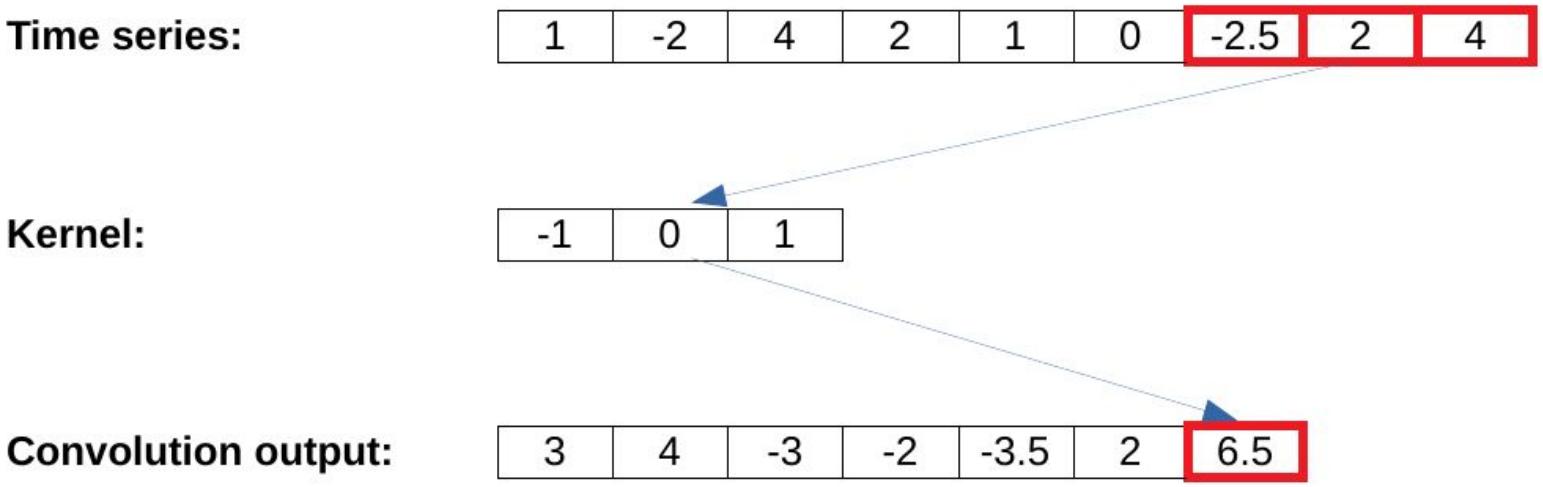
- State-of-the-art time series classifier in terms of both speed and accuracy.
- Applicable to regression problems.
- Key idea: Extract features from time series using random convolution kernels + off-the-shelf ML (RidgeCV)



ROCKET Time Series Classifier/ Regressor

Angus Dempster, Francois Petitjean, and Geoffrey I Webb - ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels - Data Mining and Knowledge Discovery / arXiv:1910.13051

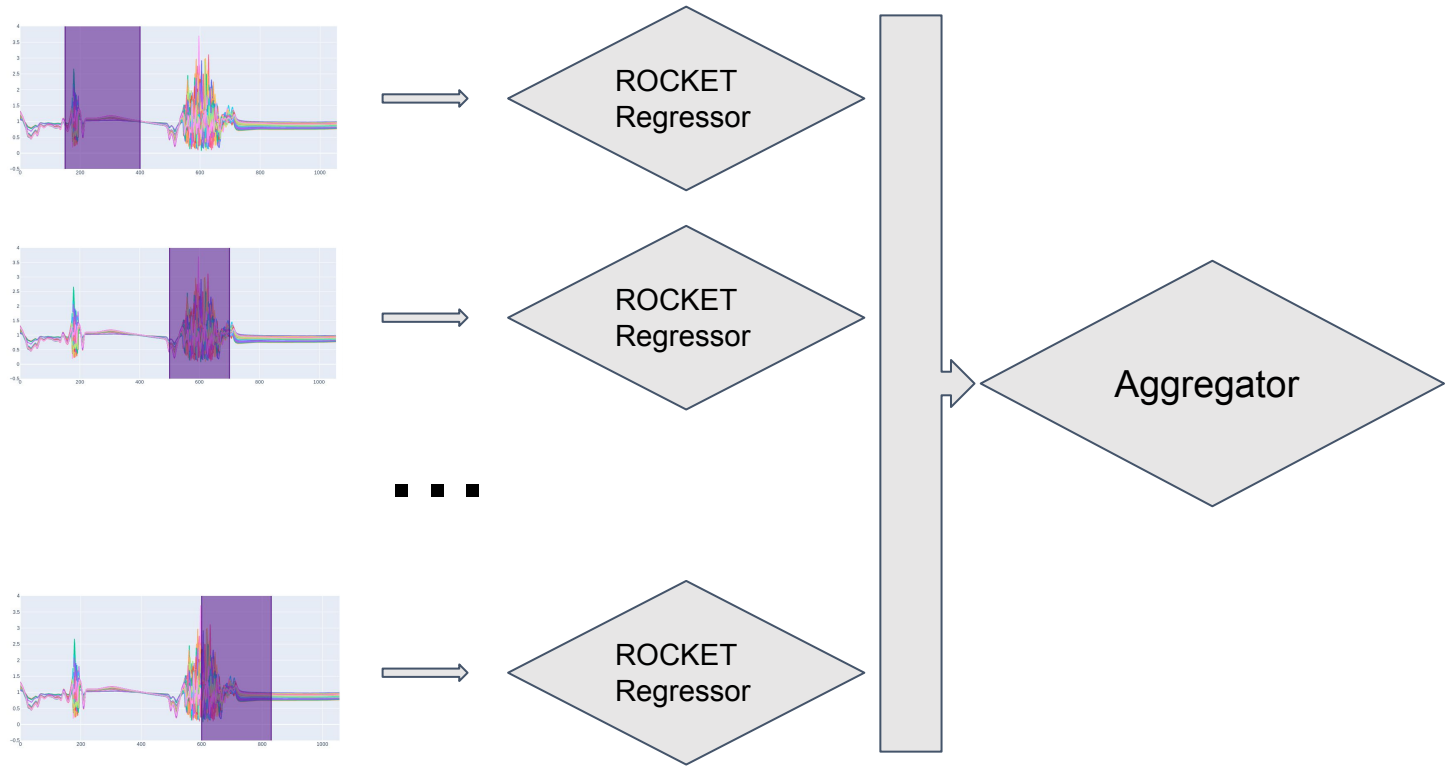
- State-of-the-art time series classifier in terms of both speed and accuracy.
- Applicable to regression problems.
- Key ideas: Extract features from time series using random convolution kernels + off-the-shelf ML (RidgeCV)



Max pooling feature: maximum value (6.5)
PPV feature: the positive ratio (4/7)

Ensemble of ROCKETS

- Our approach to improve ROCKET: Train individual ROCKET models on different intervals of the time series data.
- Aggregate the predictions of individual models



Ensemble of ROCKETS

- Train individual ROCKET models on different intervals of the time series data.
- Aggregate the predictions of individual models.

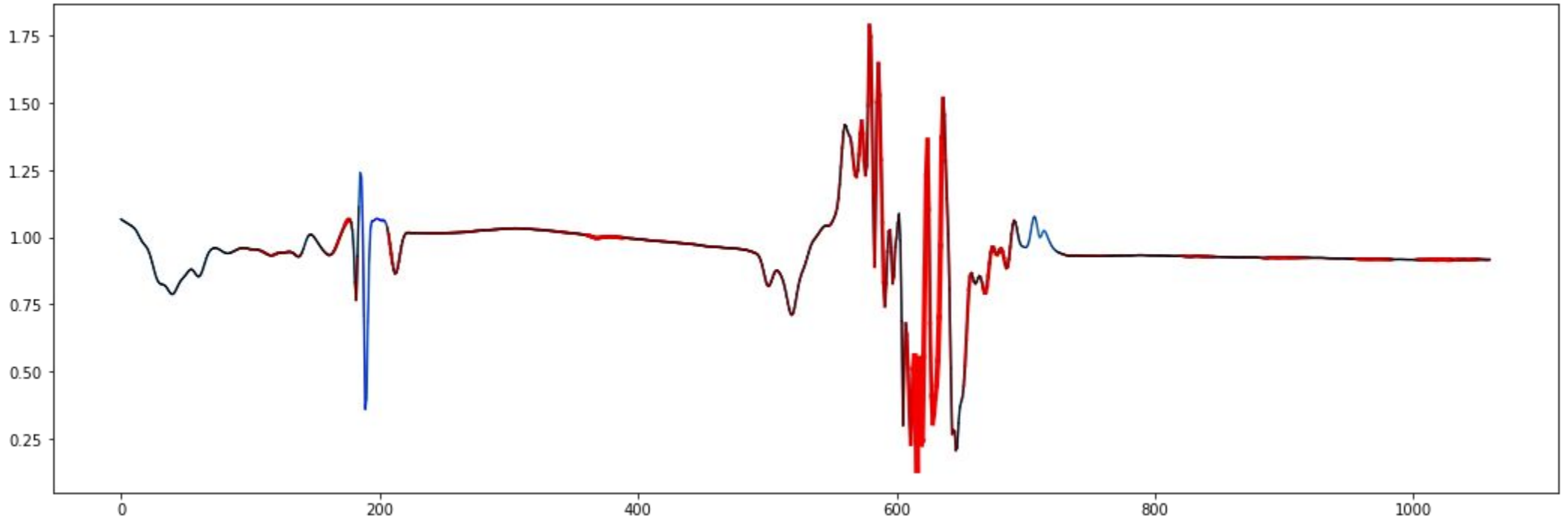
- We have explored some ensemble strategies:
  - Average of the predictions (simplest).
  - Cross validation to rank individual models.
  - Bagging (train each individual model with a subset of the training data)
  - Stacking (train a meta-regressor on top of the ensemble).
  - AdaBoost

Results

Method	Kappa-RMSECV	Micelle-RMSECV	Ph-RMSECV
LASSO(normalize=True)	1.5117	56.7817	0.1188
RidgeCV-tabular	1.1697	57.1684	0.0821
MiniROCKET	**1.1863**	**60.4678**	**0.0811**
EnsembleMiniROCKET-Avg	**1.1740**	**58.3019**	**0.0768**
EnsembleMiniROCKET-CV	**1.1648**	**56.8801**	**0.0737**

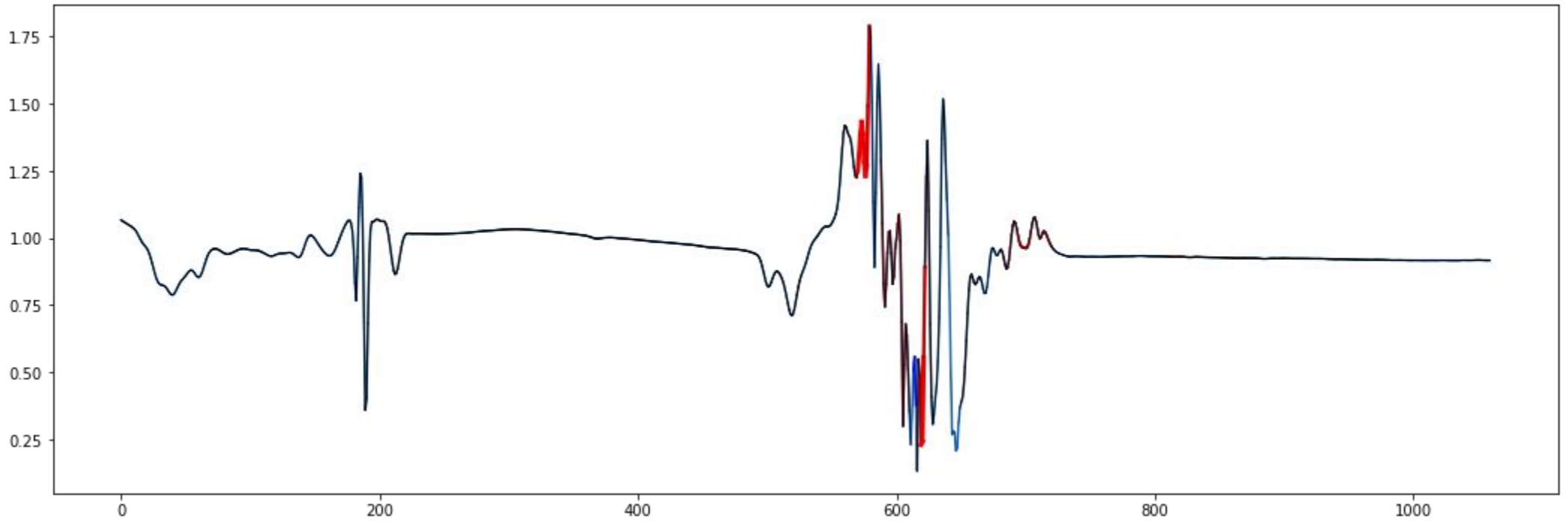
- The results are comparable if not slightly better than the tabular approaches.
- Ensemble seems to help here. However, it needs to be done carefully.
- It is still unclear to us whether the time series approaches are suitable for this particular problem.

Local Explanation with MrSQM



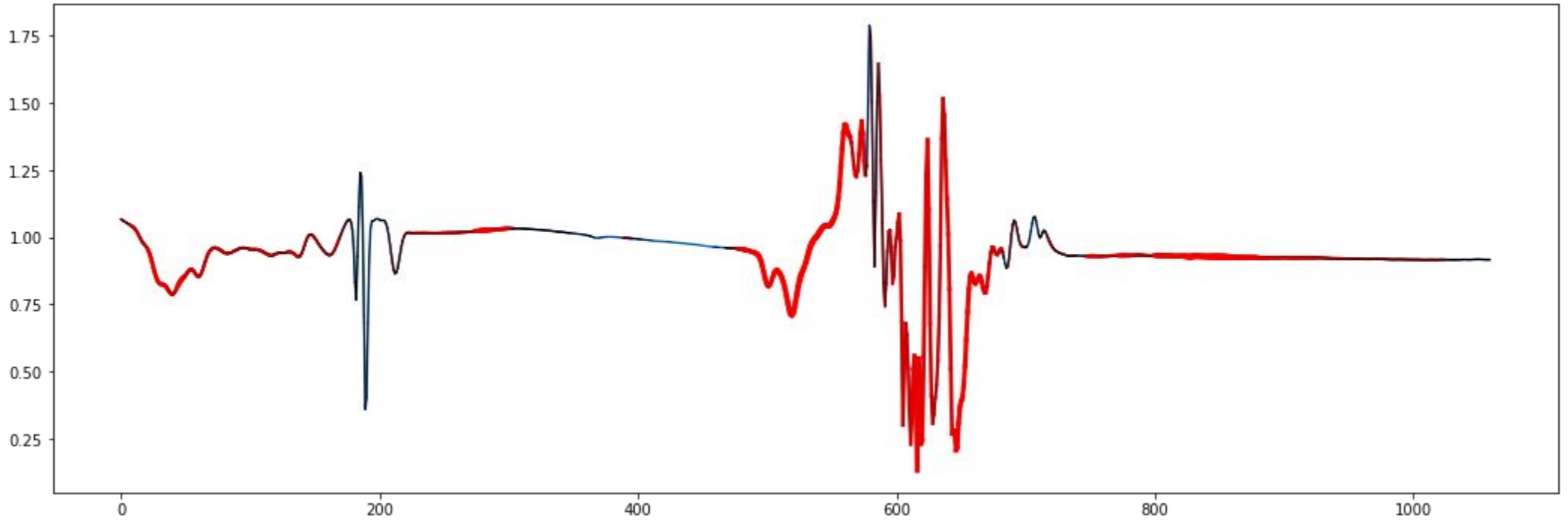
- Trait: kappa_casein
- Intercept: 4.0386
- Red increases the target value (positive coefficient), while blue decreases it (negative coefficient).

Local Explanation with MrSQM



- Trait: Casein_micelle_size
- Intercept: 186.907
- Red increases the target value while blue decreases it.

Local Explanation with MrSQM



- Trait: Native_Ph
- Intercept: 6.69
- Red increases the target value while blue decreases it.


VistaMilk
STRATEGIC GOALS
FOOD SECURITY
Health & Resilience
Safe Foods
Food Provenance
New Foods
Smart Ingredients
SUSTAINABILITY
Climate
Environment
Biodiversity
Circular Economy
Efficiency
PROSPERITY & SOCIETAL ENRICHMENT
Job & Wealth Creation
Socio-economic
Education
Engagement
Knowledge Transfer