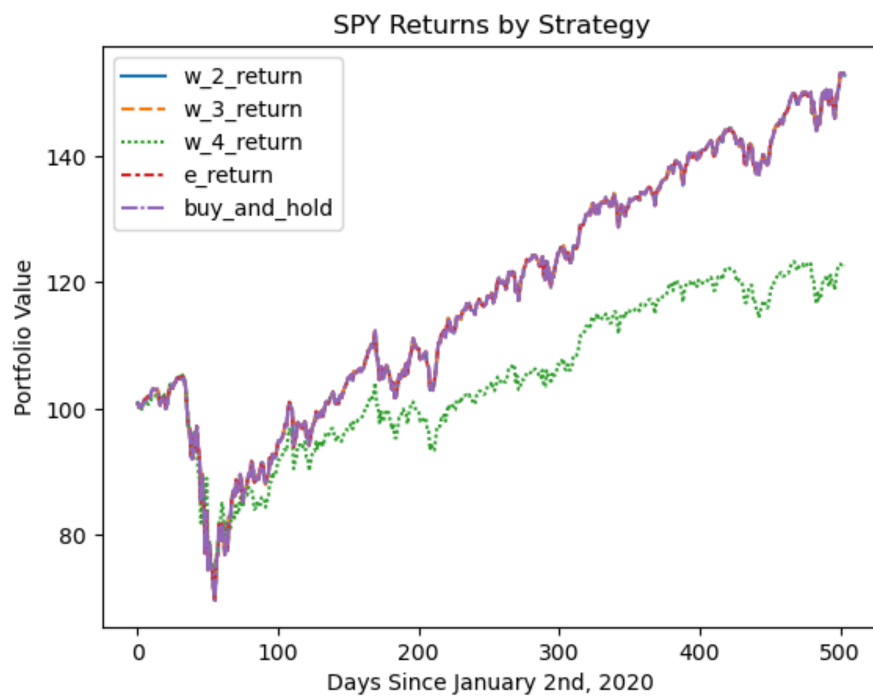
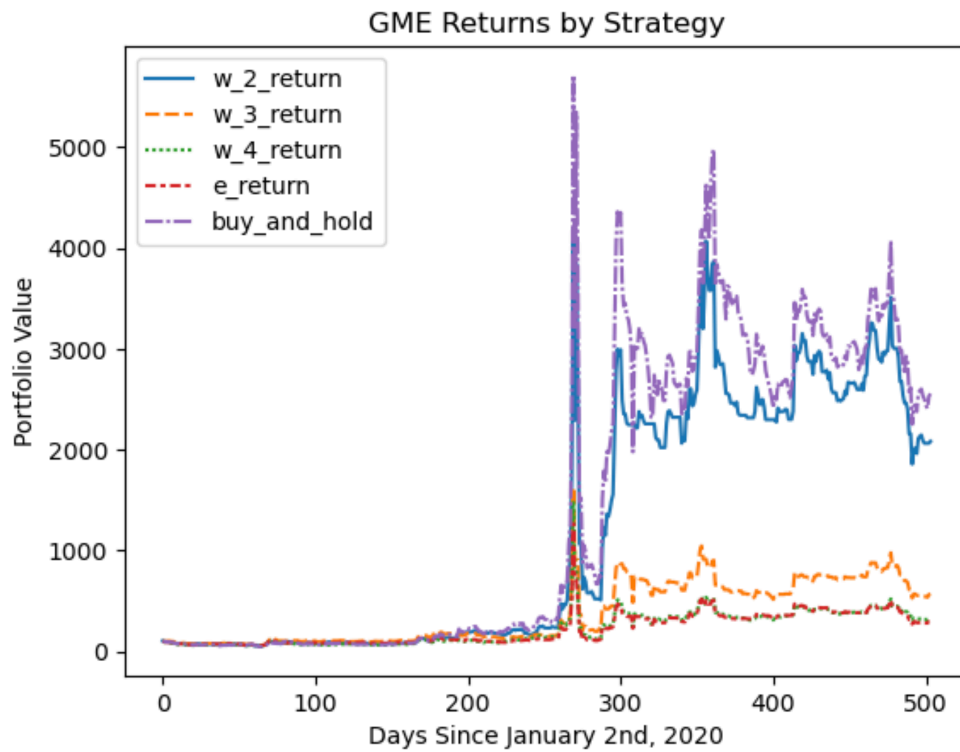


Hw 11 Will Mohr CS677A3Fall23

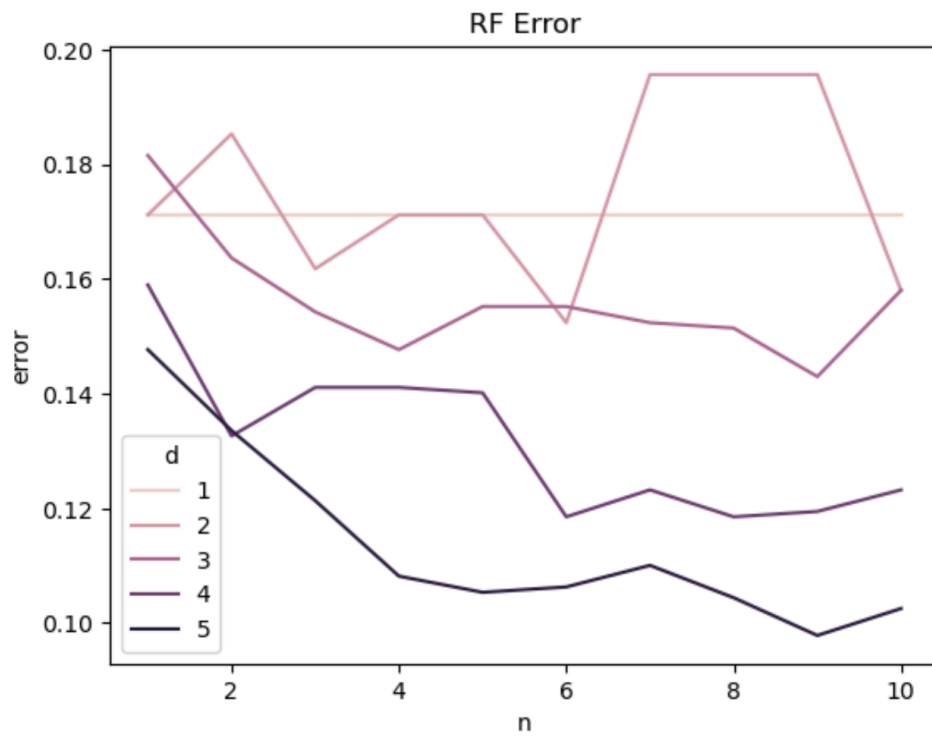
Comparisons

		TP	FP	TN	FN	accuracy	TPR	TNR	Trading	B&H
gme	lda	30	1	17	4	0.90	0.88	0.94	49767.72	807.43
	knn_5	28	1	17	6	0.87	0.82	0.94	45809.4	807.43
	svc_linear	27	1	17	7	0.85	0.79	0.94	45327.32	807.43
	knn_3	29	1	17	5	0.88	0.85	0.94	44865.61	807.43
	logistic	26	1	17	8	0.83	0.76	0.94	44369.46	807.43
	random_forest	26	0	18	8	0.85	0.76	1.00	42120.43	807.43
	decision_tree	28	1	17	6	0.87	0.82	0.94	39126.11	807.43
	knn_1	26	1	17	8	0.83	0.76	0.94	31453.2	807.43
	svc_gaussian	28	1	17	6	0.87	0.82	0.94	9172.4	807.43
	qda	26	2	16	8	0.81	0.76	0.89	8032.21	807.43
	svc_poly	29	11	7	5	0.69	0.85	0.39	2599.65	807.43
spy	logistic	19	0	31	2	0.96	0.90	1.00	165.06	127.54
	svc_linear	19	1	30	2	0.94	0.90	0.97	164.84	127.54
	lda	16	0	31	5	0.90	0.76	1.00	163.88	127.54
	knn_5	20	0	31	1	0.98	0.95	1.00	163.66	127.54
	decision_tree	19	4	27	2	0.88	0.90	0.87	162.93	127.54
	random_forest	20	2	29	1	0.94	0.95	0.94	162.87	127.54
	knn_3	21	0	31	0	1.00	1.00	1.00	162.73	127.54
	svc_gaussian	21	1	30	0	0.98	1.00	0.97	162.51	127.54
	knn_1	21	2	29	0	0.96	1.00	0.94	161.95	127.54
	qda	11	0	31	10	0.81	0.52	1.00	158.74	127.54
	svc_poly	0	0	31	21	0.60	0.00	1.00	129.05	127.54



Cardiography

	TP	FP	TN	FN	accuracy	TPR	TNR
GaussianNB	737	105	139	82	0.82	0.90	0.57
LogisticRegression	773	147	97	46	0.82	0.94	0.40
DecisionTreeClassifier	759	67	177	60	0.88	0.93	0.73
RandomForestClassifier	797	83	161	22	0.90	0.97	0.66
SVC_linear	774	147	97	45	0.82	0.95	0.40
SVC_degree2	811	200	44	8	0.80	0.99	0.18
SVC_gaussian	790	83	161	29	0.89	0.96	0.66



One way to find the importance of features would be to start from scratch. try each feature individually, proceed with the best feature per test set or validation results. Next, try adding each remaining feature to the model one at a time. proceed with the feature that improves the test set the most. then try removing each feature one at a time (one fewer features than full model) and see if any previously included features are now redundant and lead to overfitting. proceed until no additional features lead to a statistically significant improvement.

SVM — finance

		TP	FP	TN	FN	accuracy	TPR	TNR	Trading	B&H
SPY	svc_linear	19	1	30	2	0.94	0.90	0.97	164.84	127.54
	svc_gaussian	21	1	30	0	0.98	1.00	0.97	162.51	127.54
	svc_poly	0	0	31	21	0.60	0.00	1.00	129.05	127.54
GME	svc_linear	27	1	17	7	0.85	0.79	0.94	45327.32	807.43
	svc_gaussian	28	1	17	6	0.87	0.82	0.94	9172.4	807.43
	svc_poly	29	11	7	5	0.69	0.85	0.39	2599.65	807.43