

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

имени М.В.ЛОМОНОСОВА



ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

КАФЕДРА ЭКОНОМИКИ ИННОВАЦИЙ

РАЗДАТОЧНЫЙ МАТЕРИАЛ К ВКР

**«ВЗАИМОСВЯЗЬ ДИНАМИКИ ЦЕН АКЦИЙ И НЕФИНАНСОВЫХ ПОТОКОВ
ИНФОРМАЦИИ НА ПРИМЕРЕ АВТОМОБИЛЬНЫХ КОМПАНИЙ»**

The Interrelation of Share Price Dynamics and Non-Financial Information: a Case of
Automobile Companies

Выполнила: студентка группы э403

Горшкова Марина Олеговна

Научный руководитель:

Мирзоян Ашот Гамлетович

Москва

2023

Результаты проведенного анализа по компании Ferrari



Таблица 1. ARIMA-модель для модели с применением sentiment-анализа

Источник: составлено автором

	estimate	sd	t_stat	pi_val
intercept	0.0010	0.0005	1.8970	0.0578
sent_1	− 0.0004	0.0002	− 1.8482	0.0646
AIC	− 5 329			
BIC	− 5 314			
n.obs	1 007			

Как видно из таблицы, новостной sentiment статистически значим при первом его лагировании, то есть рынку для реакции на новостной фон необходим один день. Знак оценки коэффициента отрицательный, что говорит об ошибочности «мгновенной» реакции рынка на новую информацию.

Таблица 2. Прогнозирование доходностей акций компании с помощью линейной ARIMA-модели на всей выборке

Источник: составлено автором

	int_a_arima	cor_arima	int_b_arima	pv_arima
ferr	0.04944	0.14392	0.23585	0.01258

При использовании линейных моделей имеет смысл прогнозирование доходностей только на новостной информации: корреляция прогнозов с истинными значениями на выборке обучения довольно большая и статистически значимая.

Таблица 3. Прогнозирование больших по модулю доходностей с использованием sentiment-анализа

Источник: Составлено автором

	int_a_lm	cor_lm	int_b_lm	pv_lm
ferr	−0.214488	0.030091	0.271119	0.840869

Прогнозирование больших доходностей (в контексте отклонения от модуля медианы) с использованием sentiment-анализа является нецелесообразным, поскольку дает статистически незначимые результаты. Возможно, между новостным фоном и большими по модулю доходностями имеется нелинейная связь.

Таблица 4. Прогнозирование доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	- 0.0068	0.1821	0.0884	0.1265
	Poly	- 0.1966	- 0.0083	- 0.1034	0.0738
	Radial	- 0.1067	0.0835	- 0.0117	0.8399
Только сентимент в усеченной форме	Linear	- 0.1973	- 0.0090	- 0.1041	0.0719
	Poly	- 0.1417	0.0482	- 0.0471	0.4159
	Radial	- 0.0494	0.1405	0.0459	0.4280
Только сентимент в значимой форме	Linear	- 0.2358	- 0.0494	- 0.1439	0.0126
	Poly	- 0.2358	- 0.0494	- 0.1439	0.0126
	Radial	- 0.0324	0.1571	0.0629	0.2773
Финансы + усеченная форма сентимента	Linear	- 0.0917	0.0986	0.0035	0.9516
	Poly	- 0.0816	0.1086	0.0136	0.8143
	Radial	- 0.0830	0.1073	0.0123	0.8320
Финансы + значимая форма сентимента	Linear	- 0.0863	0.1040	0.0089	0.8778
	Poly	- 0.0429	0.1469	0.0525	0.3653
	Radial	- 0.0958	0.0945	- 0.0007	0.9907

Таблица 5. Прогнозирование больших доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	- 0.0068	0.2577	0.1277	0.1181
	Poly	- 0.1837	0.0844	- 0.0505	0.5379
	Radial	- 0.2002	0.0673	- 0.0677	0.4090
Только сентимент в усеченной форме	Linear	- 0.0281	0.2377	0.1067	0.1921
	Poly	- 0.0656	0.2019	0.0694	0.3971
	Radial	- 0.1124	0.1562	0.0223	0.7858
Только сентимент в значимой форме	Linear	0.1148	0.3676	0.2454	0.0024
	Poly	0.1148	0.3676	0.2454	0.0024
	Radial	- 0.1145	0.1542	0.0202	0.8051
Финансы и усеченная форма сентимента	Linear	0.0057	0.2693	0.1400	0.0864
	Poly	- 0.0332	0.2329	0.1017	0.2142
	Radial	- 0.0561	0.2111	0.0789	0.3354
Финансы и значимая форма сентимента	Linear	0.1511	0.3992	0.2798	0.0005
	Poly	0.0741	0.3316	0.2064	0.0110
	Radial	- 0.2179	0.0489	- 0.0861	0.2934

Использование ML-методов в прогнозировании всех доходностей при сентимент-анализе нецелесообразно. При больших доходностях корреляция прогнозов высокая и значимая, что говорит об обратном.

Таблица 6. Классификация по направлению движения акций с использованием методов машинного обучения и sentiment-анализа

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4936	0.5452	0.5194	-
	Random_forest	0.3887	0.5668	0.4777	-
	Bayes_Naive	0.4767	0.5381	0.5074	-
	Boosting	0.4595	0.5295	0.4945	-
	SVM-Linear	0.5183	0.5304	0.5244	+
	SVM-Poly	0.5107	0.5380	0.5244	+
	SVM-Radial	0.4547	0.5364	0.4956	-
Использование только новостной информации (текущая и первый лаг)	Logit	0.5004	0.5485	0.5244	+
	Random_forest	0.4666	0.5763	0.5214	-
	Bayes_Naive	0.4445	0.5226	0.4835	-
	Boosting	0.4600	0.5829	0.5215	-
	SVM-Linear	0.5050	0.5378	0.5214	+
	SVM-Poly	0.5183	0.5304	0.5244	+
	SVM-Radial	0.4709	0.5700	0.5205	-
Использование лагированной новостной информации	Logit	0.4601	0.5689	0.5145	-
	Random_forest	0.4758	0.5769	0.5264	-
	Bayes_Naive	0.4160	0.5528	0.4844	-
	Boosting	0.4583	0.5646	0.5115	-
	SVM-Linear	0.5038	0.5351	0.5194	+
	SVM-Poly	0.5183	0.5304	0.5244	+
	SVM-Radial	0.4858	0.6107	0.5482	-
Оба канала информации (текущая и первый лаг)	Logit	0.4743	0.5666	0.5204	-
	Random_forest	0.4093	0.5522	0.4807	-
	Bayes_Naive	0.4468	0.5243	0.4855	-
	Boosting	0.4449	0.5521	0.4985	-
	SVM-Linear	0.5050	0.5378	0.5214	+
	SVM-Poly	0.5183	0.5304	0.5244	+
	SVM-Radial	0.4413	0.5777	0.5095	-
Оба канала информации (новости лагированные)	Logit	0.4466	0.5604	0.5035	-
	Random_forest	0.4039	0.5555	0.4797	-
	Bayes_Naive	0.3991	0.5620	0.4805	-
	Boosting	0.4351	0.5262	0.4806	-
	SVM-Linear	0.5038	0.5351	0.5194	+
	SVM-Poly	0.5183	0.5304	0.5244	+
	SVM-Radial	0.4285	0.5943	0.5114	-

В любой спецификации модели, построенные с помощью метода опорных векторов и линейного или полиномиального ядра статистически значимо лучше, чем слепой классификатор. Точность моделей вне зависимости от источника данных сопоставима.

Таблица 7. Классификация по размаху движения акций с использованием методов машинного обучения и sentiment-анализа

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4584	0.5863	0.5223	-
	Random_forest	0.4198	0.5652	0.4925	-
	Bayes_Naive	0.4829	0.6015	0.5422	-
	Boosting	0.5121	0.5645	0.5383	+
	SVM-Linear	0.4572	0.5438	0.5005	-
	SVM-Poly	0.4979	0.5807	0.5393	-
	SVM-Radial	0.4429	0.6180	0.5304	-
Использование только новостной информации (текущая и первый лаг)	Logit	0.4464	0.5724	0.5094	-
	Random_forest	0.4587	0.5562	0.5074	-
	Bayes_Naive	0.4795	0.5574	0.5184	-
	Boosting	0.4361	0.5766	0.5064	-
	SVM-Linear	0.4630	0.5478	0.5054	-
	SVM-Poly	0.4789	0.5360	0.5074	-
	SVM-Radial	0.4516	0.5692	0.5104	-
Использование лагированной новостной информации	Logit	0.4770	0.5616	0.5193	-
	Random_forest	0.4047	0.5566	0.4806	-
	Bayes_Naive	0.4234	0.5616	0.4925	-
	Boosting	0.4364	0.5586	0.4975	-
	SVM-Linear	0.4508	0.5381	0.4945	-
	SVM-Poly	0.4588	0.5601	0.5094	-
	SVM-Radial	0.4555	0.5516	0.5036	-
Оба канала информации (текущая и первый лаг)	Logit	0.4420	0.5808	0.5114	-
	Random_forest	0.4281	0.5608	0.4944	-
	Bayes_Naive	0.4874	0.6090	0.5482	-
	Boosting	0.4738	0.5967	0.5353	-
	SVM-Linear	0.4691	0.5656	0.5173	-
	SVM-Poly	0.4988	0.5877	0.5432	-
	SVM-Radial	0.4824	0.5962	0.5393	-
Оба канала информации (новости лагированные)	Logit	0.4406	0.5683	0.5045	-
	Random_forest	0.4325	0.5645	0.4985	-
	Bayes_Naive	0.4284	0.5904	0.5094	-
	Boosting	0.4476	0.5711	0.5093	-
	SVM-Linear	0.4607	0.5363	0.4985	-
	SVM-Poly	0.4869	0.5936	0.5402	-
	SVM-Radial	0.4631	0.5898	0.5264	-

В случае компании Феррари прогнозирование «аномальных» доходностей представляется возможным только в одном случае – при применении бустингового алгоритма на финансовых данных. Возможно, это объясняется особенностями компании.

Таблица 8. ARIMA-модель для Феррари с выделением топиков по методу LDA

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ar1	– 0.9135	0.0420	– 21.7512	0.0000
ma1	0.9563	0.0307	31.1549	0.0000
intercept	0.0016	0.0006	2.4719	0.0134
topik_5_1	– 0.0202	0.0085	– 2.3806	0.0173
topik_6_2	0.0105	0.0055	1.8977	0.0577
topik_4_2	– 0.0077	0.0042	– 1.8202	0.0687
topik_1_3	– 0.0093	0.0056	– 1.6702	0.0949
AIC	– 5 339			
BIC	– 5 300			
n.obs	1 007			

Применительно к компании Феррари статистически значимыми являются тематические группы, относящиеся к компании как к бренду; затрагивающие вопросы производства компаний-конкурентов, о технических характеристиках автомобилей Феррари, а также о новостях касательно Формулы-1, гоночного соревнования, в котором команда Феррари принимает участие.

Таблица 9. Прогнозы доходностей с помощью модели с использованием LDA

Источник: составлено автором

	int_a_arima_dir	int_b_arima_dir	cor_arima_dir	pv_arima_dir
ferr	0.0630	0.2826	0.1750	0.0024

Прогнозы, построенные с использованием словарей, составленных по методу Латентного размещения Дирихле, имеют статистически значимую корреляцию с истинными значениями доходностей Феррари. Причем по своему среднему значению она превышает аналогичный показатель, полученный при оценивании ARIMA-модели с использованием сентимент-анализа.

Таблица 10. Прогнозы дневных доходностей с помощью методов машинного обучения и словарей, составленных с применением LDA

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Финансовая информация	Linear	-0.0716	0.1544	0.0419	0.4693
	Poly	-0.2216	0.0021	-0.1112	0.0544
	Radial	-0.0739	0.1523	0.0397	0.4932

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только текущие словари	Linear	-0.1148	0.1117	-0.0016	0.9783
	Poly	-0.1460	0.0803	-0.0333	0.5660
	Radial	-0.1819	0.0435	-0.0701	0.2260
Только лаги словарей	Linear	-0.0813	0.1449	0.0322	0.5786
	Poly	-0.0743	0.1519	0.0393	0.4977
	Radial	-0.0485	0.1770	0.0651	0.2611
2 источника – текущие словари	Linear	-0.1126	0.1138	0.0006	0.9916
	Poly	-0.1355	0.0909	-0.0226	0.6965
	Radial	-0.0812	0.1450	0.0323	0.5774
2 источника – лаги словарей	Linear	-0.0632	0.1627	0.0504	0.3846
	Poly	-0.0640	0.1619	0.0496	0.3922
	Radial	-0.1737	0.0519	-0.0617	0.2869

Использование методов машинного обучения при прогнозировании значений доходности компании Феррари не приносит статистически значимых результатов и при использовании тематических словарей в качестве регрессоров.

Таблица 11. Классификация направления изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4936	0.5452	0.5194	-
	Random_forest	0.3887	0.5668	0.4777	-
	Bayes_Naive	0.4767	0.5381	0.5074	-
	Boosting	0.4595	0.5295	0.4945	-
	SVM-Linear	0.5183	0.5304	0.5244	+
	SVM-Poly	0.5107	0.5380	0.5244	+
	SVM-Radial	0.4547	0.5364	0.4956	-
Использование только новостной информации (текущая)	Logit	0.4924	0.5744	0.5334	-
	Random_forest	0.4543	0.5886	0.5214	-
	Bayes_Naive	0.4503	0.5369	0.4936	-
	Boosting	0.4792	0.5716	0.5254	-
	SVM-Linear	0.4854	0.5753	0.5304	-
	SVM-Poly	0.4896	0.5731	0.5314	-
	SVM-Radial	0.4768	0.5899	0.5334	-
Использование лагированной новостной информации	Logit	0.4511	0.5999	0.5255	-
	Random_forest	0.4434	0.5876	0.5155	-
	Bayes_Naive	0.4113	0.6019	0.5066	-
	Boosting	0.4776	0.6051	0.5414	-
	SVM-Linear	0.4356	0.5915	0.5136	-
	SVM-Poly	0.4642	0.5947	0.5294	-
	SVM-Radial	0.4326	0.6005	0.5165	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (текущая)	Logit	0.4861	0.5747	0.5304	-
	Random_forest	0.4326	0.5845	0.5085	-
	Bayes_Naive	0.4548	0.5165	0.4856	-
	Boosting	0.4367	0.5603	0.4985	-
	SVM-Linear	0.4854	0.5753	0.5304	-
	SVM-Poly	0.4896	0.5731	0.5314	-
	SVM-Radial	0.4966	0.5900	0.5433	-
Оба канала информации (новости лагированные)	Logit	0.4324	0.5986	0.5155	-
	Random_forest	0.4620	0.5530	0.5075	-
	Bayes_Naive	0.4102	0.5990	0.5046	-
	Boosting	0.4508	0.6079	0.5293	-
	SVM-Linear	0.4377	0.5933	0.5155	-
	SVM-Poly	0.4900	0.5687	0.5294	-
	SVM-Radial	0.4657	0.6028	0.5342	-

Таблица 12. Классификация магнитуды изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4584	0.5863	0.5223	-
	Random_forest	0.4198	0.5652	0.4925	-
	Bayes_Naive	0.4829	0.6015	0.5422	-
	Boosting	0.5121	0.5645	0.5383	+
	SVM-Linear	0.4572	0.5438	0.5005	-
	SVM-Poly	0.4979	0.5807	0.5393	-
	SVM-Radial	0.4429	0.6180	0.5304	-
Использование только новостной информации (текущая)	Logit	0.4433	0.5617	0.5025	-
	Random_forest	0.4527	0.5722	0.5125	-
	Bayes_Naive	0.4274	0.5837	0.5055	-
	Boosting	0.4496	0.5634	0.5065	-
	SVM-Linear	0.4364	0.5608	0.4986	-
	SVM-Poly	0.4759	0.5530	0.5144	-
	SVM-Radial	0.4367	0.5624	0.4996	-
Использование лагированной новостной информации	Logit	0.4296	0.5613	0.4954	-
	Random_forest	0.4802	0.5605	0.5204	-
	Bayes_Naive	0.4242	0.5669	0.4955	-
	Boosting	0.4647	0.5621	0.5134	-
	SVM-Linear	0.4095	0.5595	0.4845	-
	SVM-Poly	0.4522	0.5488	0.5005	-
	SVM-Radial	0.4184	0.5587	0.4886	-
Оба канала информации (текущая)	Logit	0.4556	0.5754	0.5155	-
	Random_forest	0.4813	0.5855	0.5334	-
	Bayes_Naive	0.4695	0.5892	0.5293	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (новости лагированные)	Boosting	0.4727	0.6078	0.5402	-
	SVM-Linear	0.4362	0.5730	0.5046	-
	SVM-Poly	0.5128	0.5877	0.5502	+
	SVM-Radial	0.4640	0.5850	0.5245	-
	Logit	0.3853	0.6097	0.4975	-
	Random_forest	0.4174	0.6232	0.5203	-
	Bayes_Naive	0.4300	0.5551	0.4926	-
	Boosting	0.4826	0.6099	0.5462	-
	SVM-Linear	0.4255	0.5395	0.4825	-
	SVM-Poly	0.4450	0.5540	0.4995	-
	SVM-Radial	0.4286	0.6061	0.5173	-

Для решения задач классификации применение LDA-словарей не делает модели лучше по качеству, чем случайное угадывание.

Таблица 13. Классификация направления изменения доходности по важным для компании Феррари словам (TF-IDF)

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4936	0.5452	0.5194	-
	Random_forest	0.3887	0.5668	0.4777	-
	Bayes_Naive	0.4767	0.5381	0.5074	-
	Boosting	0.4595	0.5295	0.4945	-
	SVM-Linear	0.5183	0.5304	0.5244	+
	SVM-Poly	0.5107	0.5380	0.5244	+
	SVM-Radial	0.4547	0.5364	0.4956	-
Использование только новостной информации (текущая)	Logit	0.5015	0.5553	0.5284	+
	Random_forest	0.5040	0.5468	0.5254	+
	Bayes_Naive	0.4936	0.5492	0.5214	-
	Boosting	0.5183	0.5304	0.5244	+
	SVM-Linear	0.5004	0.5483	0.5244	+
	SVM-Poly	0.5102	0.5466	0.5284	+
	SVM-Radial	0.5183	0.5304	0.5244	+
Использование лагированной новостной информации	Logit	0.5059	0.5559	0.5309	+
	Random_forest	0.5044	0.5533	0.5289	+
	Bayes_Naive	0.4781	0.5638	0.5210	-
	Boosting	0.5175	0.5303	0.5239	+
	SVM-Linear	0.5025	0.5533	0.5279	+
	SVM-Poly	0.5075	0.5443	0.5259	+
	SVM-Radial	0.4989	0.5649	0.5319	-
Оба канала информации (текущая)	Logit	0.4829	0.5640	0.5234	-
	Random_forest	0.5183	0.5304	0.5244	+

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (новости лагированные)	Bayes_Naive	0.4891	0.5477	0.5184	-
	Boosting	0.4595	0.5295	0.4945	-
	SVM-Linear	0.5027	0.5481	0.5254	+
	SVM-Poly	0.5102	0.5466	0.5284	+
	SVM-Radial	0.5007	0.5521	0.5264	+
	Logit	0.4896	0.5683	0.5289	-
	Random_forest	0.5175	0.5303	0.5239	+
	Bayes_Naive	0.4781	0.5638	0.5210	-
	Boosting	0.4132	0.5710	0.4921	-
	SVM-Linear	0.5039	0.5539	0.5289	+
	SVM-Poly	0.5075	0.5443	0.5259	+
	SVM-Radial	0.4770	0.5807	0.5289	-

Как модели, использующие только новостную информацию, так и комбинирующие 2 источника данных показывают точность выше, чем случайный классификатор. При добавлении в них новостной информации качество классификации улучшается.

Таблица 14. Классификация масштаба изменения доходности по важным для компании Феррари словам (TF-IDF)

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4584	0.5863	0.5223	-
	Random_forest	0.4198	0.5652	0.4925	-
	Bayes_Naive	0.4829	0.6015	0.5422	-
	Boosting	0.5121	0.5645	0.5383	+
	SVM-Linear	0.4572	0.5438	0.5005	-
	SVM-Poly	0.4979	0.5807	0.5393	-
	SVM-Radial	0.4429	0.6180	0.5304	-
Использование только новостной информации (текущая)	Logit	0.4756	0.5374	0.5065	-
	Random_forest	0.4768	0.5382	0.5075	-
	Bayes_Naive	0.4888	0.5201	0.5045	-
	Boosting	0.4935	0.5055	0.4995	-
	SVM-Linear	0.4873	0.5296	0.5085	-
	SVM-Poly	0.4906	0.5303	0.5104	-
	SVM-Radial	0.4860	0.5310	0.5085	-
Использование лагированной новостной информации	Logit	0.4727	0.5193	0.4960	-
	Random_forest	0.4799	0.5280	0.5040	-
	Bayes_Naive	0.4827	0.5214	0.5020	-
	Boosting	0.4939	0.5042	0.4990	-
	SVM-Linear	0.4779	0.5181	0.4980	-
	SVM-Poly	0.4860	0.5260	0.5060	-
	SVM-Radial	0.4917	0.5103	0.5010	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (текущая)	Logit	0.4660	0.5669	0.5165	-
	Random_forest	0.4786	0.5503	0.5144	-
	Bayes_Naive	0.4931	0.5198	0.5065	-
	Boosting	0.5121	0.5645	0.5383	+
	SVM-Linear	0.4738	0.5312	0.5025	-
	SVM-Poly	0.4709	0.5599	0.5154	-
	SVM-Radial	0.4547	0.5702	0.5124	-
Оба канала информации (новости ланированные)	Logit	0.4456	0.5823	0.5140	-
	Random_forest	0.4684	0.5796	0.5240	-
	Bayes_Naive	0.4827	0.5214	0.5020	-
	Boosting	0.4862	0.6135	0.5498	-
	SVM-Linear	0.4474	0.5308	0.4891	-
	SVM-Poly	0.4836	0.5422	0.5129	-
	SVM-Radial	0.4669	0.5410	0.5040	-

Модели классификации амплитуды колебаний дневных доходностей компании Феррари с использованием данных о частоте употребления «важных» для компании слов справляется с задачей не лучше слепого классификатора.



Таблица 1. ARIMA-модель для модели с применением сентимент-анализа

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ar1	− 0.5850	0.1299	− 4.5024	0.0000
ma1	0.7843	0.1312	5.9797	0.0000
ma2	0.1230	0.0493	2.4933	0.0127
ma3	0.0221	0.0417	0.5290	0.5968
ma4	− 0.0804	0.0364	− 2.2054	0.0274
intercept	0.0004	0.0009	0.4557	0.6486
sent_7	0.0003	0.0001	2.1425	0.0322
sent_2	− 0.0002	0.0001	− 1.9178	0.0551
sent	0.0002	0.0001	1.7820	0.0748
AIC	− 4 638			
BIC	− 4 589			
n.obs	1 007			

Таблица 2. Прогнозирование доходностей акций компании с помощью линейной ARIMA-модели на всей выборке

Источник: составлено автором

	int_a_arima	int_b_arima	cor_arima	pv_arima
ford	− 0.1142	0.0761	− 0.0192	0.7404

Таблица 3. Прогнозирование больших по модулю доходностей с использованием сентимент-анализа

Источник: Составлено автором

	int_a_lm	int_b_lm	cor_lm	pv_lm
ford	− 0.1775	0.0908	− 0.0441	0.5904

В результате оценивания линейной ARIMA-модели статистически значимо на дневные доходности влияют текущий сентимент, а также его второй и седьмой лаги. Однако прогнозирование на всей выборке только по текстовой информации не имеет в себе статистически значимой корреляции прогнозов с истинными значениями, равно как и аналогичная процедура, но для больших по модулю доходностей.

Таблица 4. Прогнозирование доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	0.1841	0.3602	0.2745	0.0000
	Poly	0.1778	0.3545	0.2684	0.0000
	Radial	0.1693	0.3468	0.2602	0.0000
Только сентимент в усеченной форме	Linear	-0.1917	-0.0032	-0.0984	0.0890
	Poly	-0.1424	0.0474	-0.0479	0.4082
	Radial	-0.0221	0.1672	0.0732	0.2060
Только сентимент в значимой форме	Linear	-0.1197	0.0705	-0.0249	0.6680
	Poly	-0.0685	0.1217	0.0268	0.6435
	Radial	-0.1156	0.0747	-0.0206	0.7217
Финансы + усеченная форма сентимента	Linear	0.1591	0.3376	0.2504	0.0000
	Poly	0.1359	0.3164	0.2281	0.0001
	Radial	0.1092	0.2918	0.2022	0.0004
Финансы + значимая форма сентимента	Linear	0.1570	0.3357	0.2485	0.0000
	Poly	0.0543	0.2404	0.1487	0.0099
	Radial	0.0846	0.2689	0.1783	0.0019

Таблица 5. Прогнозирование больших доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	0.2405	0.4744	0.3632	0.0000
	Poly	0.2387	0.4729	0.3615	0.0000
	Radial	0.2198	0.4572	0.3440	0.0000
Только сентимент в усеченной форме	Linear	-0.2536	0.0112	-0.1234	0.1313
	Poly	-0.1687	0.0997	-0.0351	0.6685
	Radial	-0.0734	0.1944	0.0616	0.4522
Только сентимент в значимой форме	Linear	-0.2200	0.0467	-0.0882	0.2814
	Poly	-0.2160	0.0510	-0.0840	0.3051
	Radial	-0.1546	0.1141	-0.0206	0.8014
Финансы и усеченная форма сентимента	Linear	0.1372	0.3871	0.2666	0.0009
	Poly	0.0626	0.3213	0.1953	0.0162
	Radial	0.1388	0.3886	0.2682	0.0009
Финансы и значимая форма сентимента	Linear	0.1947	0.4363	0.3207	0.0001
	Poly	0.1472	0.3958	0.2761	0.0006
	Radial	0.0903	0.3461	0.2220	0.0062

Включение финансовой информации в модели прогнозирования дневной доходности компании Форд, применяющие методы машинного обучения не ведет к увеличению их качества, измеряемого в данном случае корреляцией прогнозов с истинными значениями.

Таблица 6. Классификация по направлению движения акций с использованием методов машинного обучения и sentiment-анализа

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5034	0.6787	0.5910	+
	Random_forest	0.4863	0.6124	0.5494	-
	Bayes_Naive	0.4750	0.6833	0.5791	-
	Boosting	0.4558	0.6848	0.5703	-
	SVM-Linear	0.5071	0.6451	0.5761	+
	SVM-Poly	0.5138	0.6484	0.5811	+
	SVM-Radial	0.5060	0.6922	0.5991	+
Использование только новостной информации (текущая и первый лаг)	Logit	0.4654	0.5496	0.5075	-
	Random_forest	0.4188	0.5424	0.4806	-
	Bayes_Naive	0.4587	0.5363	0.4975	-
	Boosting	0.4265	0.5645	0.4955	-
	SVM-Linear	0.5084	0.5205	0.5144	+
	SVM-Poly	0.5084	0.5205	0.5144	+
	SVM-Radial	0.4396	0.5553	0.4975	-
Использование лагированной новостной информации	Logit	0.4565	0.6161	0.5363	-
	Random_forest	0.4547	0.5922	0.5235	-
	Bayes_Naive	0.4376	0.5692	0.5034	-
	Boosting	0.4005	0.5727	0.4866	-
	SVM-Linear	0.4896	0.5373	0.5134	-
	SVM-Poly	0.4838	0.5689	0.5264	-
	SVM-Radial	0.4351	0.6079	0.5215	-
Оба канала информации (текущая и первый лаг)	Logit	0.5201	0.6441	0.5821	+
	Random_forest	0.5077	0.6528	0.5802	+
	Bayes_Naive	0.4789	0.6514	0.5651	-
	Boosting	0.4535	0.6829	0.5682	-
	SVM-Linear	0.5260	0.6163	0.5711	+
	SVM-Poly	0.4859	0.6782	0.5821	-
	SVM-Radial	0.4766	0.6678	0.5722	-
Оба канала информации (новости лагированные)	Logit	0.4958	0.6683	0.5821	-
	Random_forest	0.4726	0.6699	0.5712	-
	Bayes_Naive	0.4652	0.6351	0.5502	-
	Boosting	0.4804	0.6620	0.5712	-
	SVM-Linear	0.4805	0.6617	0.5711	-
	SVM-Poly	0.4826	0.6596	0.5711	-
	SVM-Radial	0.4593	0.6672	0.5633	-

Таблица 7. Классификация по размаху движения акций с использованием методов машинного обучения и sentiment-анализа

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4259	0.5812	0.5036	-
	Random_forest	0.5005	0.6516	0.5761	+
	Bayes_Naive	0.5790	0.6628	0.6209	+
	Boosting	0.5679	0.6581	0.6130	+
	SVM-Linear	0.4974	0.5555	0.5264	-
	SVM-Poly	0.5403	0.5900	0.5651	+
	SVM-Radial	0.5568	0.6988	0.6278	+
Использование только новостной информации (текущая и первый лаг)	Logit	0.4710	0.5280	0.4995	-
	Random_forest	0.4494	0.5635	0.5064	-
	Bayes_Naive	0.4821	0.5587	0.5204	-
	Boosting	0.4607	0.5723	0.5165	-
	SVM-Linear	0.4771	0.5258	0.5015	-
	SVM-Poly	0.4804	0.5445	0.5124	-
	SVM-Radial	0.4685	0.5782	0.5233	-
Использование лагированной новостной информации	Logit	0.4229	0.5545	0.4887	-
	Random_forest	0.4058	0.6114	0.5086	-
	Bayes_Naive	0.4680	0.5588	0.5134	-
	Boosting	0.4235	0.5598	0.4916	-
	SVM-Linear	0.4340	0.5613	0.4976	-
	SVM-Poly	0.4717	0.5374	0.5045	-
	SVM-Radial	0.4475	0.5794	0.5135	-
Оба канала информации (текущая и первый лаг)	Logit	0.4534	0.5895	0.5215	-
	Random_forest	0.4979	0.6383	0.5681	-
	Bayes_Naive	0.5343	0.6058	0.5701	+
	Boosting	0.5516	0.6465	0.5991	+
	SVM-Linear	0.4689	0.5460	0.5074	-
	SVM-Poly	0.5264	0.6079	0.5672	+
	SVM-Radial	0.5291	0.6827	0.6059	+
Оба канала информации (новости лагированные)	Logit	0.4246	0.5905	0.5075	-
	Random_forest	0.5072	0.6510	0.5791	+
	Bayes_Naive	0.5089	0.6055	0.5572	+
	Boosting	0.5489	0.6612	0.6050	+
	SVM-Linear	0.4574	0.5577	0.5075	-
	SVM-Poly	0.5245	0.5940	0.5592	+
	SVM-Radial	0.5203	0.6817	0.6010	+

Таблица 8. ARIMA-модель для Форд с выделением топигов по методу LDA

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ar1	− 0.5919	0.1262	− 4.6910	0.0000
ma1	0.7901	0.1280	6.1722	0.0000
ma2	0.1333	0.0493	2.7027	0.0069
ma3	0.0449	0.0425	1.0558	0.2911
ma4	− 0.0654	0.0373	− 1.7524	0.0797
intercept	0.0020	0.0013	1.5156	0.1296
topik_9_1	− 0.0165	0.0066	− 2.4880	0.0129
topik_9_2	− 0.0161	0.0066	− 2.4209	0.0155
topik_10_3	0.0208	0.0122	1.7035	0.0885
topik_8_2	− 0.0169	0.0097	− 1.7390	0.0820
topik_1_3	− 0.0095	0.0046	− 2.0582	0.0396
AIC	− 4 647			
BIC	− 4 578			
n.obs	1 007			

Статистически значимыми тематическими разделами являются: новости, посвященные общей динамике на рынке (их первые два лага), второй лаг новостей, посвященных нефти, а также третьи лаги таких тем как динамика финансовых показателей и новости касательно моделей автомобилей марки Ford.

Таблица 9. Прогнозы доходностей с помощью модели с использованием LDA

Источник: составлено автором

	int_a_arima_dir	int_b_arima_dir	cor_arima_dir	pv_arima_dir
ford	0.0074	0.2306	0.1205	0.0370

Наблюдается статистически значимая корреляция прогнозов модели, построенной на принадлежности новостей к тематическим группам, и истинных значений дневной доходности компании Форд

Таблица 10. Прогнозы дневных доходностей с помощью ML и LDA-словарей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Финансовая информация	Linear	0.1732	0.3821	0.2810	0.0000
	Poly	0.1736	0.3824	0.2813	0.0000
	Radial	0.0507	0.2712	0.1630	0.0047
Только текущие словари	Linear	-0.1769	0.0487	-0.0649	0.2622
	Poly	-0.1885	0.0366	-0.0769	0.1840

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только лаги словарей	Radial	-0.2449	-0.0226	-0.1354	0.0189
	Linear	-0.0205	0.2041	0.0930	0.1080
	Poly	-0.0369	0.1883	0.0767	0.1853
	Radial	-0.0117	0.2125	0.1017	0.0787
2 источника – текущие словари	Linear	0.1565	0.3672	0.2650	0.0000
	Poly	0.0975	0.3142	0.2084	0.0003
	Radial	0.0043	0.2277	0.1174	0.0421
2 источника – лаги словарей	Linear	0.1852	0.3926	0.2923	0.0000
	Poly	0.1569	0.3677	0.2655	0.0000
	Radial	0.0908	0.3081	0.2019	0.0004

Модели, построенные только на лагированной информации словарей, имеют статистически значимую корреляцию при использовании радиально-базисного ядра в методе опорных векторов, также добавление этого потока информации положительно сказывается на качестве моделей, комбинирующих источники данных.

Таблица 11. Классификация направления изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5034	0.6787	0.5910	+
	Random_forest	0.4863	0.6124	0.5494	-
	Bayes_Naive	0.4750	0.6833	0.5791	-
	Boosting	0.4558	0.6848	0.5703	-
	SVM-Linear	0.5071	0.6451	0.5761	+
	SVM-Poly	0.5138	0.6484	0.5811	+
	SVM-Radial	0.5060	0.6922	0.5991	+
Использование только новостной информации (текущая)	Logit	0.4392	0.6096	0.5244	-
	Random_forest	0.4804	0.5545	0.5175	-
	Bayes_Naive	0.4280	0.6011	0.5145	-
	Boosting	0.4667	0.5923	0.5295	-
	SVM-Linear	0.4316	0.6213	0.5264	-
	SVM-Poly	0.4314	0.6235	0.5274	-
	SVM-Radial	0.4296	0.6253	0.5274	-
Использование лагированной новостной информации	Logit	0.4214	0.6052	0.5133	-
	Random_forest	0.4223	0.5647	0.4935	-
	Bayes_Naive	0.4087	0.5981	0.5034	-
	Boosting	0.4489	0.5620	0.5054	-
	SVM-Linear	0.5087	0.5242	0.5164	+
	SVM-Poly	0.4393	0.5913	0.5153	-
	SVM-Radial	0.4214	0.6052	0.5133	-
	Logit	0.5218	0.6504	0.5861	+

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (текущая)	Random_forest	0.4696	0.6629	0.5663	-
	Bayes_Naive	0.4944	0.5861	0.5403	-
	Boosting	0.4731	0.6814	0.5772	-
	SVM-Linear	0.5047	0.6336	0.5692	+
	SVM-Poly	0.5130	0.6433	0.5781	+
	SVM-Radial	0.4726	0.6679	0.5702	-
Оба канала информации (новости лагированные)	Logit	0.4731	0.6792	0.5762	-
	Random_forest	0.4772	0.6014	0.5393	-
	Bayes_Naive	0.4284	0.5887	0.5085	-
	Boosting	0.4878	0.7082	0.5980	-
	SVM-Linear	0.4551	0.6255	0.5403	-
	SVM-Poly	0.4762	0.6363	0.5563	-
	SVM-Radial	0.4712	0.6133	0.5422	-

Таблица 12. Классификация магнитуды изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4259	0.5812	0.5036	-
	Random_forest	0.5005	0.6516	0.5761	+
	Bayes_Naive	0.5790	0.6628	0.6209	+
	Boosting	0.5679	0.6581	0.6130	+
	SVM-Linear	0.4974	0.5555	0.5264	-
	SVM-Poly	0.5403	0.5900	0.5651	+
	SVM-Radial	0.5568	0.6988	0.6278	+
Использование только новостной информации (текущая)	Logit	0.4977	0.6030	0.5503	-
	Random_forest	0.4599	0.6050	0.5324	-
	Bayes_Naive	0.4669	0.6099	0.5384	-
	Boosting	0.4997	0.5969	0.5483	-
	SVM-Linear	0.4914	0.6013	0.5463	-
	SVM-Poly	0.4911	0.6076	0.5493	-
	SVM-Radial	0.5122	0.5944	0.5533	+
Использование лагированной новостной информации	Logit	0.4242	0.6226	0.5234	-
	Random_forest	0.4775	0.6328	0.5552	-
	Bayes_Naive	0.4524	0.5982	0.5253	-
	Boosting	0.4486	0.6418	0.5452	-
	SVM-Linear	0.4069	0.5842	0.4956	-
	SVM-Poly	0.4450	0.6414	0.5432	-
	SVM-Radial	0.4264	0.6540	0.5402	-
Оба канала информации (текущая)	Logit	0.5088	0.5977	0.5533	+
	Random_forest	0.5365	0.6495	0.5930	+
	Bayes_Naive	0.5297	0.6424	0.5860	+
	Boosting	0.5245	0.6896	0.6071	+
	SVM-Linear	0.4936	0.6011	0.5473	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
	SVM-Poly	0.5153	0.5991	0.5572	+
	SVM-Radial	0.5478	0.6481	0.5980	+
Оба канала информации (новости лагированные)	Logit	0.4247	0.6002	0.5125	-
	Random_forest	0.5299	0.6660	0.5979	+
	Bayes_Naive	0.4874	0.6209	0.5542	-
	Boosting	0.5308	0.6771	0.6039	+
	SVM-Linear	0.4011	0.5940	0.4975	-
	SVM-Poly	0.4847	0.6634	0.5740	-
	SVM-Radial	0.4505	0.6916	0.5711	-

Применение информации, полученной на основании тематического распределения новостей в большинстве своем имеет смысл только при прогнозировании числовых значений доходности.

Таблица 13. Классификация направления изменения доходности по важным для компании Форд словам (TF-IDF)

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5034	0.6787	0.5910	+
	Random_forest	0.4863	0.6124	0.5494	-
	Bayes_Naive	0.4750	0.6833	0.5791	-
	Boosting	0.4558	0.6848	0.5703	-
	SVM-Linear	0.5071	0.6451	0.5761	+
	SVM-Poly	0.5138	0.6484	0.5811	+
	SVM-Radial	0.5060	0.6922	0.5991	+
Использование только новостной информации (текущая)	Logit	0.4879	0.5529	0.5204	-
	Random_forest	0.4926	0.5461	0.5194	-
	Bayes_Naive	0.5007	0.5301	0.5154	+
	Boosting	0.4886	0.5423	0.5154	-
	SVM-Linear	0.4983	0.5424	0.5204	-
	SVM-Poly	0.5068	0.5340	0.5204	+
	SVM-Radial	0.5084	0.5205	0.5144	+
Использование лагированной новостной информации	Logit	0.4716	0.5344	0.5030	-
	Random_forest	0.5075	0.5204	0.5139	+
	Bayes_Naive	0.4832	0.5247	0.5040	-
	Boosting	0.4764	0.5375	0.5069	-
	SVM-Linear	0.4713	0.5207	0.4960	-
	SVM-Poly	0.5075	0.5204	0.5139	+
	SVM-Radial	0.5075	0.5204	0.5139	+
Оба канала информации (текущая)	Logit	0.5093	0.6786	0.5940	+
	Random_forest	0.5175	0.6208	0.5692	+
	Bayes_Naive	0.5019	0.5269	0.5144	+

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (новости лагированные)	Boosting	0.4711	0.7071	0.5891	-
	SVM-Linear	0.4954	0.5434	0.5194	-
	SVM-Poly	0.5068	0.5340	0.5204	+
	SVM-Radial	0.5144	0.6459	0.5802	+
	Logit	0.5489	0.6465	0.5977	+
	Random_forest	0.5074	0.6581	0.5827	+
	Bayes_Naive	0.4857	0.5442	0.5150	-
	Boosting	0.5246	0.6568	0.5907	+
	SVM-Linear	0.4574	0.5823	0.5199	-
	SVM-Poly	0.5075	0.5204	0.5139	+
	SVM-Radial	0.5039	0.6554	0.5797	+

При добавлении лагированной информации о тематическом распределении новостей в «финансовые» модели машинного обучения, их точность увеличивается.

Таблица 14. Классификация масштаба доходности по важным для компании словам

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4259	0.5812	0.5036	-
	Random_forest	0.5005	0.6516	0.5761	+
	Bayes_Naive	0.5790	0.6628	0.6209	+
	Boosting	0.5679	0.6581	0.6130	+
	SVM-Linear	0.4974	0.5555	0.5264	-
	SVM-Poly	0.5403	0.5900	0.5651	+
	SVM-Radial	0.5568	0.6988	0.6278	+
Использование только новостной информации (текущая)	Logit	0.4727	0.5640	0.5184	-
	Random_forest	0.4618	0.5809	0.5213	-
	Bayes_Naive	0.4889	0.5220	0.5055	-
	Boosting	0.4779	0.5589	0.5184	-
	SVM-Linear	0.4561	0.5547	0.5054	-
	SVM-Poly	0.4767	0.5561	0.5164	-
	SVM-Radial	0.4669	0.5698	0.5184	-
Использование лагированной новостной информации	Logit	0.4669	0.5698	0.5184	-
	Random_forest	0.4669	0.5429	0.5049	-
	Bayes_Naive	0.4580	0.5401	0.4990	-
	Boosting	0.4777	0.5381	0.5079	-
	SVM-Linear	0.4741	0.5397	0.5069	-
	SVM-Poly	0.4638	0.5163	0.4900	-
	SVM-Radial	0.4905	0.5235	0.5070	-
Оба канала информации (текущая)	Logit	0.4834	0.5891	0.5363	-
	Random_forest	0.5425	0.6455	0.5940	+
	Bayes_Naive	0.4592	0.5755	0.5174	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
	Boosting	0.5653	0.6407	0.6030	+
	SVM-Linear	0.4553	0.5556	0.5054	-
	SVM-Poly	0.4767	0.5561	0.5164	-
	SVM-Radial	0.5200	0.6817	0.6009	+
Оба канала информации (новости ланированные)	Logit	0.4195	0.6240	0.5218	-
	Random_forest	0.5222	0.6650	0.5936	+
	Bayes_Naive	0.4758	0.5520	0.5139	-
	Boosting	0.5047	0.7202	0.6125	+
	SVM-Linear	0.4659	0.5222	0.4940	-
	SVM-Poly	0.4918	0.5261	0.5090	-
	SVM-Radial	0.5269	0.6862	0.6065	+



Таблица 1. ARIMA-модель для модели с применением сентимент-анализа

Источник: составлено автором

	estimate	sd	t_stat	pi_val
intercept	- 0.0009	0.0007	- 1.3051	0.1919
sent_4	0.0006	0.0003	1.7432	0.0813
sent_5	0.0006	0.0003	1.7225	0.0850
AIC	- 4 848			
BIC	- 4 829			
n.obs	1 007			

Таблица 2. Прогнозирование доходностей акций компании с помощью ARIMA-модели на всей выборке

Источник: составлено автором

	int_a_arima	int_b_arima	cor_arima	pv_arima
niss	- 0.0025	0.1862	0.0927	0.1090

Таблица 3. Прогнозирование больших по модулю доходностей с использованием сентимент-анализа

Источник: Составлено автором

	int_a_lm	int_b_lm	cor_lm	pv_lm
niss	-0.0941	0.1743	0.0408	0.6187

Таблица 4. Прогнозирование доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	-0.1134	0.0769	-0.0184	0.7509
	Poly	-0.1496	0.0401	-0.0553	0.3402
	Radial	-0.1675	0.0218	-0.0735	0.2041
Только сентимент в усеченной форме	Linear	-0.0452	0.1446	0.0502	0.3865
	Poly	-0.0630	0.1272	0.0324	0.5763
	Radial	-0.1142	0.0761	-0.0192	0.7399
Только сентимент в значимой форме	Linear	-0.0056	0.1832	0.0896	0.1216
	Poly	-0.0031	0.1856	0.0921	0.1114
	Radial	-0.1061	0.0842	-0.0111	0.8484
Финансы + усеченная форма сентимента	Linear	-0.0456	0.1442	0.0498	0.3905
	Poly	-0.0631	0.1270	0.0323	0.5778
	Radial	-0.2002	-0.0120	-0.1071	0.0641

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Финансы + значимая форма сентимента	Linear	-0.0031	0.1857	0.0921	0.1113
	Poly	-0.0025	0.1862	0.0927	0.1090
	Radial	-0.1813	0.0076	-0.0876	0.1299

Таблица 5. Прогнозирование больших доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	-0.1365	0.1323	-0.0022	0.9789
	Poly	-0.1884	0.0796	-0.0554	0.4995
	Radial	-0.1311	0.1377	0.0033	0.9676
Только сентимент в усеченной форме	Linear	-0.0220	0.2434	0.1127	0.1681
	Poly	-0.0084	0.2562	0.1262	0.1226
	Radial	-0.2092	0.0580	-0.0769	0.3477
Только сентимент в значимой форме	Linear	-0.0463	0.2204	0.0886	0.2792
	Poly	-0.0438	0.2228	0.0912	0.2657
	Radial	-0.2367	0.0292	-0.1056	0.1967
Финансы и усеченная форма сентимента	Linear	-0.0243	0.2412	0.1104	0.1771
	Poly	-0.0108	0.2540	0.1238	0.1299
	Radial	-0.1847	0.0833	-0.0516	0.5289
Финансы и значимая форма сентимента	Linear	-0.0482	0.2186	0.0868	0.2895
	Poly	-0.0490	0.2178	0.0859	0.2940
	Radial	-0.1470	0.1218	-0.0128	0.8758

Таблица 6. Классификация по направлению движения акций с использованием методов машинного обучения и сентимент-анализа

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4796	0.5392	0.5094	-
	Random_forest	0.4314	0.5757	0.5035	-
	Bayes_Naive	0.4748	0.5879	0.5313	-
	Boosting	0.4503	0.5725	0.5114	-
	SVM-Linear	0.5147	0.5241	0.5194	+
	SVM-Poly	0.4996	0.5610	0.5303	-
	SVM-Radial	0.4194	0.5838	0.5016	-
Использование только новостной информации (текущая и первый лаг)	Logit	0.4950	0.5497	0.5224	-
	Random_forest	0.4706	0.5661	0.5184	-
	Bayes_Naive	0.4716	0.5594	0.5155	-
	Boosting	0.4689	0.5500	0.5094	-
	SVM-Linear	0.5147	0.5241	0.5194	+
	SVM-Poly	0.5147	0.5241	0.5194	+
	SVM-Radial	0.5074	0.5434	0.5254	+

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование лагированной новостной информации	Logit	0.4773	0.5775	0.5274	-
	Random_forest	0.4725	0.5841	0.5283	-
	Bayes_Naive	0.4891	0.5874	0.5383	-
	Boosting	0.4640	0.5766	0.5203	-
	SVM-Linear	0.5043	0.5346	0.5194	+
	SVM-Poly	0.5072	0.5554	0.5313	+
	SVM-Radial	0.4768	0.5758	0.5263	-
Оба канала информации (текущая и первый лаг)	Logit	0.4897	0.5490	0.5194	-
	Random_forest	0.4130	0.5842	0.4986	-
	Bayes_Naive	0.4403	0.6022	0.5213	-
	Boosting	0.3986	0.6345	0.5166	-
	SVM-Linear	0.5147	0.5241	0.5194	+
	SVM-Poly	0.4839	0.5886	0.5362	-
	SVM-Radial	0.4597	0.5811	0.5204	-
Оба канала информации (новости лагированные)	Logit	0.4764	0.5585	0.5175	-
	Random_forest	0.4131	0.5918	0.5025	-
	Bayes_Naive	0.4762	0.5964	0.5363	-
	Boosting	0.4507	0.6021	0.5264	-
	SVM-Linear	0.5074	0.5394	0.5234	+
	SVM-Poly	0.4925	0.5781	0.5353	-
	SVM-Radial	0.4635	0.5434	0.5034	-

Использование информации о сентименте новостей для компании Ниссан ведет к повышению точности моделей, отличных по качеству от случайного угадывания при их сравнении с моделями, построенными только на текстовых данных (или же классификации на том же уровне).

Таблица 7. Классификация по размаху движения акций с использованием методов машинного обучения и сентимент-анализа

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.3880	0.5595	0.4737	-
	Random_forest	0.4575	0.6233	0.5404	-
	Bayes_Naive	0.4792	0.7011	0.5901	-
	Boosting	0.4843	0.6620	0.5732	-
	SVM-Linear	0.4019	0.5475	0.4747	-
	SVM-Poly	0.4864	0.5605	0.5234	-
	SVM-Radial	0.4849	0.6835	0.5842	-
Использование только новостной информации	Logit	0.4859	0.5648	0.5253	-
	Random_forest	0.4835	0.5452	0.5144	-
	Bayes_Naive	0.4762	0.5466	0.5114	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
(текущая и первый лаг)	Boosting	0.4794	0.5493	0.5144	-
	SVM-Linear	0.4768	0.5441	0.5104	-
	SVM-Poly	0.4855	0.5593	0.5224	-
	SVM-Radial	0.4934	0.5573	0.5254	-
Использование лагированной новостной информации	Logit	0.4303	0.6042	0.5172	-
	Random_forest	0.4172	0.6096	0.5134	-
	Bayes_Naive	0.4782	0.5566	0.5174	-
	Boosting	0.4562	0.5786	0.5174	-
	SVM-Linear	0.4596	0.5889	0.5243	-
	SVM-Poly	0.5058	0.5569	0.5313	+
	SVM-Radial	0.4345	0.5644	0.4995	-
Оба канала информации (текущая и первый лаг)	Logit	0.3934	0.6018	0.4976	-
	Random_forest	0.4979	0.6067	0.5523	-
	Bayes_Naive	0.4901	0.6144	0.5522	-
	Boosting	0.4857	0.6846	0.5852	-
	SVM-Linear	0.4239	0.5970	0.5104	-
	SVM-Poly	0.4801	0.5727	0.5264	-
	SVM-Radial	0.5070	0.6453	0.5762	+
Оба канала информации (новости лагированные)	Logit	0.4094	0.5935	0.5014	-
	Random_forest	0.5172	0.6450	0.5811	+
	Bayes_Naive	0.4721	0.6085	0.5403	-
	Boosting	0.5099	0.6543	0.5821	+
	SVM-Linear	0.4550	0.5698	0.5124	-
	SVM-Poly	0.5149	0.5876	0.5512	+
	SVM-Radial	0.4904	0.6498	0.5701	-

Комбинация источников данных ведет к наилучшим результатам классификации дневных доходностей по факту их отклонения от модуля медианного значения.

Таблица 8. ARIMA-модель для Ниссан с выделением топиков по методу LDA

Источник: составлено автором

	estimate	sd	t_stat	pi_val
intercept	0.0009	0.0009	0.9950	0.3197
topik_3_3	0.0102	0.0063	1.6327	0.1025
topik_10	- 0.0268	0.0105	- 2.5498	0.0108
topik_4_3	0.0330	0.0163	2.0218	0.0432
topik_3_2	- 0.0151	0.0061	- 2.4533	0.0142
topik_4_2	- 0.0332	0.0157	- 2.1168	0.0343
topik_10_3	- 0.0246	0.0117	- 2.1091	0.0349
AIC	-4,862			
BIC	-4,823			
n.obs	1,007			

Статистически значимой является информация, получаемая из следующих тематических групп: относительно поставок комплектующих (причем это единственный показатель, на который рынок реагирует моментально), модельного ряда (с двумя порядками легирования), а также нвоости, посвященные административным вопросам относительно компании.

Таблица 9. Прогнозы доходностей с помощью модели с использованием LDA

Источник: составлено автором

	int_a_arima_dir	int_b_arima_dir	cor_arima_dir	pv_arima_dir
niss	0.0255	0.2477	0.1383	0.0165

Таблица 10. Прогнозы дневных доходностей с помощью ML и LDA-словарей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Финансовая информация	Linear	-0.2009	0.0238	-0.0897	0.1211
	Poly	-0.1999	0.0248	-0.0887	0.1254
	Radial	-0.2255	-0.0020	-0.1152	0.0462
Только текущие словари	Linear	-0.1423	0.0840	-0.0296	0.6100
	Poly	-0.1186	0.1079	-0.0054	0.9254
	Radial	-0.1505	0.0757	-0.0379	0.5134
Только лаги словарей	Linear	-0.1571	0.0690	-0.0446	0.4414
	Poly	-0.1466	0.0796	-0.0339	0.5581
	Radial	-0.1091	0.1174	0.0042	0.9424
2 источника – текущие словари	Linear	-0.1635	0.0624	-0.0512	0.3770
	Poly	-0.1263	0.1001	-0.0133	0.8191
	Radial	-0.0932	0.1332	0.0202	0.7270
2 источника – лаги словарей	Linear	-0.1698	0.0559	-0.0577	0.3195
	Poly	-0.1493	0.0769	-0.0367	0.5266
	Radial	-0.1109	0.1155	0.0023	0.9680

Таблица 11. Классификация направления изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4796	0.5392	0.5094	-
	Random_forest	0.4314	0.5757	0.5035	-
	Bayes_Naive	0.4748	0.5879	0.5313	-
	Boosting	0.4503	0.5725	0.5114	-
	SVM-Linear	0.5147	0.5241	0.5194	+
	SVM-Poly	0.4996	0.5610	0.5303	-
	SVM-Radial	0.4194	0.5838	0.5016	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только новостной информации (текущая)	Logit	0.4594	0.5494	0.5044	-
	Random_forest	0.4456	0.5591	0.5024	-
	Bayes_Naive	0.4297	0.5414	0.4856	-
	Boosting	0.4513	0.5496	0.5004	-
	SVM-Linear	0.4738	0.5371	0.5054	-
	SVM-Poly	0.5147	0.5241	0.5194	+
	SVM-Radial	0.5147	0.5241	0.5194	+
Использование лагированной новостной информации	Logit	0.4557	0.5973	0.5265	-
	Random_forest	0.4649	0.5741	0.5195	-
	Bayes_Naive	0.4186	0.6102	0.5144	-
	Boosting	0.4477	0.5914	0.5195	-
	SVM-Linear	0.4332	0.5880	0.5106	-
	SVM-Poly	0.5147	0.5241	0.5194	+
	SVM-Radial	0.4505	0.6044	0.5275	-
Оба канала информации (текущая)	Logit	0.4468	0.5343	0.4905	-
	Random_forest	0.4457	0.5791	0.5124	-
	Bayes_Naive	0.4306	0.5385	0.4846	-
	Boosting	0.3855	0.5877	0.4866	-
	SVM-Linear	0.4692	0.5476	0.5084	-
	SVM-Poly	0.4344	0.6085	0.5214	-
	SVM-Radial	0.4824	0.6061	0.5443	-
Оба канала информации (новости лагированные)	Logit	0.4426	0.5965	0.5196	-
	Random_forest	0.4255	0.6058	0.5156	-
	Bayes_Naive	0.4112	0.6215	0.5164	-
	Boosting	0.4082	0.5671	0.4877	-
	SVM-Linear	0.4337	0.5795	0.5066	-
	SVM-Poly	0.5147	0.5241	0.5194	+
	SVM-Radial	0.4390	0.6300	0.5345	-

Таблица 12. Классификация магнитуды изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.3880	0.5595	0.4737	-
	Random_forest	0.4575	0.6233	0.5404	-
	Bayes_Naive	0.4792	0.7011	0.5901	-
	Boosting	0.4843	0.6620	0.5732	-
	SVM-Linear	0.4019	0.5475	0.4747	-
	SVM-Poly	0.4864	0.5605	0.5234	-
	SVM-Radial	0.4849	0.6835	0.5842	-
Использование только новостной информации (текущая)	Logit	0.4849	0.6835	0.5842	-
	Random_forest	0.4718	0.5788	0.5253	-
	Bayes_Naive	0.4650	0.5499	0.5074	-
	Boosting	0.4634	0.5595	0.5115	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование лагированной новостной информации	SVM-Linear	0.4759	0.5709	0.5234	-
	SVM-Poly	0.4543	0.5826	0.5184	-
	SVM-Radial	0.4634	0.5934	0.5284	-
	Logit	0.4187	0.5903	0.5045	-
	Random_forest	0.4295	0.5737	0.5016	-
	Bayes_Naive	0.4451	0.5699	0.5075	-
	Boosting	0.4255	0.5477	0.4866	-
	SVM-Linear	0.4453	0.5576	0.5015	-
	SVM-Poly	0.4762	0.5745	0.5253	-
	SVM-Radial	0.4615	0.5356	0.4986	-
Оба канала информации (текущая)	Logit	0.4603	0.5744	0.5173	-
	Random_forest	0.4609	0.6437	0.5523	-
	Bayes_Naive	0.4528	0.5941	0.5235	-
	Boosting	0.4808	0.6573	0.5690	-
	SVM-Linear	0.4522	0.5827	0.5174	-
	SVM-Poly	0.4613	0.5776	0.5195	-
	SVM-Radial	0.4835	0.6270	0.5553	-
Оба канала информации (новости лагированные)	Logit	0.4334	0.5736	0.5035	-
	Random_forest	0.4737	0.6290	0.5513	-
	Bayes_Naive	0.4465	0.5903	0.5184	-
	Boosting	0.4707	0.6537	0.5622	-
	SVM-Linear	0.4398	0.5690	0.5044	-
	SVM-Poly	0.4713	0.5994	0.5354	-
	SVM-Radial	0.4591	0.5836	0.5214	-

Таблица 13. Классификация направления изменения доходности по важным для компании Ниссан словам (TF-IDF)

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4796	0.5392	0.5094	-
	Random_forest	0.4314	0.5757	0.5035	-
	Bayes_Naive	0.4748	0.5879	0.5313	-
	Boosting	0.4503	0.5725	0.5114	-
	SVM-Linear	0.5147	0.5241	0.5194	+
	SVM-Poly	0.4996	0.5610	0.5303	-
	SVM-Radial	0.4194	0.5838	0.5016	-
Использование только новостной информации (текущая)	Logit	0.4950	0.5498	0.5224	-
	Random_forest	0.5147	0.5241	0.5194	+
	Bayes_Naive	0.5026	0.5263	0.5144	+
	Boosting	0.5147	0.5241	0.5194	+
	SVM-Linear	0.4976	0.5372	0.5174	-
	SVM-Poly	0.5087	0.5321	0.5204	+

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование лагированной новостной информации	SVM-Radial	0.5147	0.5241	0.5194	+
	Logit	0.5005	0.5374	0.5189	+
	Random_forest	0.5161	0.5238	0.5199	+
	Bayes_Naive	0.4787	0.5411	0.5099	-
	Boosting	0.5161	0.5238	0.5199	+
	SVM-Linear	0.4938	0.5360	0.5149	-
	SVM-Poly	0.5096	0.5342	0.5219	+
	SVM-Radial	0.5161	0.5238	0.5199	+
Оба канала информации (текущая)	Logit	0.4549	0.5916	0.5233	-
	Random_forest	0.5147	0.5241	0.5194	+
	Bayes_Naive	0.5026	0.5263	0.5144	+
	Boosting	0.4503	0.5725	0.5114	-
	SVM-Linear	0.4992	0.5395	0.5194	-
	SVM-Poly	0.5087	0.5321	0.5204	+
	SVM-Radial	0.4558	0.5969	0.5263	-
Оба канала информации (новости лагированные)	Logit	0.4700	0.5659	0.5180	-
	Random_forest	0.5161	0.5238	0.5199	+
	Bayes_Naive	0.4788	0.5391	0.5089	-
	Boosting	0.4756	0.5903	0.5330	-
	SVM-Linear	0.4964	0.5355	0.5159	-
	SVM-Poly	0.5096	0.5342	0.5219	+
	SVM-Radial	0.4489	0.5372	0.4931	-

Таблица 14. Классификация масштаба доходности по важным для компании словам

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.3880	0.5595	0.4737	-
	Random_forest	0.4575	0.6233	0.5404	-
	Bayes_Naive	0.4792	0.7011	0.5901	-
	Boosting	0.4843	0.6620	0.5732	-
	SVM-Linear	0.4019	0.5475	0.4747	-
	SVM-Poly	0.4864	0.5605	0.5234	-
	SVM-Radial	0.4849	0.6835	0.5842	-
Использование только новостной информации (текущая)	Logit	0.4571	0.5717	0.5144	-
	Random_forest	0.4525	0.5684	0.5104	-
	Bayes_Naive	0.4935	0.5055	0.4995	-
	Boosting	0.4851	0.5119	0.4985	-
	SVM-Linear	0.4608	0.5621	0.5114	-
	SVM-Poly	0.4713	0.5595	0.5154	-
	SVM-Radial	0.4616	0.5633	0.5125	-
	Logit	0.4779	0.5479	0.5129	-
	Random_forest	0.4750	0.5529	0.5139	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование лагированной новостной информации	Bayes_Naive	0.4820	0.5339	0.5079	-
	Boosting	0.4910	0.5189	0.5050	-
	SVM-Linear	0.4788	0.5371	0.5079	-
	SVM-Poly	0.4872	0.5367	0.5119	-
	SVM-Radial	0.4915	0.5562	0.5239	-
Оба канала информации (текущая)	Logit	0.4574	0.5555	0.5065	-
	Random_forest	0.4793	0.6552	0.5673	-
	Bayes_Naive	0.4935	0.5055	0.4995	-
	Boosting	0.4932	0.6571	0.5752	-
	SVM-Linear	0.4608	0.5621	0.5114	-
	SVM-Poly	0.4713	0.5595	0.5154	-
	SVM-Radial	0.4682	0.6741	0.5711	-
Оба канала информации (новости ланированные)	Logit	0.4723	0.5614	0.5169	-
	Random_forest	0.4554	0.6242	0.5398	-
	Bayes_Naive	0.4820	0.5339	0.5079	-
	Boosting	0.5027	0.6606	0.5817	+
	SVM-Linear	0.4807	0.5452	0.5129	-
	SVM-Poly	0.4872	0.5367	0.5119	-
	SVM-Radial	0.5016	0.6636	0.5826	+

Результаты проведенного анализа по компании Kia

Таблица 1. ARIMA-модель для модели с применением сентимент-анализа

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ar1	0.0275	0.0516	0.5335	0.5937
ar2	− 0.8960	0.0602	− 14.8778	0.0000
ar3	0.2060	0.0363	5.6809	0.0000
ar4	0.0021	0.0375	0.0551	0.9560
ma1	0.2327	0.0406	5.7332	0.0000
ma2	0.9226	0.0569	16.2233	0.0000
intercept	0.0009	0.0008	1.1135	0.2655
sent	− 0.0005	0.0002	− 2.6176	0.0089
sent_1	− 0.0005	0.0002	− 2.2236	0.0262
sent_4	0.0006	0.0002	2.7813	0.0054
AIC	− 5 031			
BIC	− 4 977			
n.obs	985			

Таблица 2. Прогнозирование доходностей акций компании с помощью ARIMA-модели на всей выборке

Источник: составлено автором

	int_a_arima	int_b_arima	cor_arima	pv_arima
kia	−0.1135	0.0842	−0.0148	0.8066

Таблица 3. Прогнозирование больших по модулю доходностей с использованием сентимент-анализа

Источник: Составлено автором

	int_a_lm	int_b_lm	cor_lm	pv_lm
kia	−0.1713	−0.1007	−0.0360	0.6654

Таблица 4. Прогнозирование доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	−0.1134	0.0769	−0.0184	0.7509
	Poly	−0.1096	0.0881	−0.0109	0.8564
	Radial	−0.1675	0.0218	−0.0735	0.2041
Только сентимент в усеченной форме	Linear	−0.0452	0.1446	0.0502	0.3865
	Poly	−0.0630	0.1272	0.0324	0.5763
	Radial	−0.0731	0.1245	0.0260	0.6665

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только sentiment в значимой форме	Linear	-0.0056	0.1832	0.0896	0.1216
	Poly	-0.0031	0.1856	0.0921	0.1114
	Radial	-0.1061	0.0842	-0.0111	0.8484
Финансы + усеченная форма sentiment	Linear	0.0949	0.2854	0.1919	0.0013
	Poly	-0.0111	0.1851	0.0879	0.1439
	Radial	-0.0512	0.1461	0.0479	0.4264
Финансы + значимая форма sentiment	Linear	0.0944	0.2850	0.1915	0.0013
	Poly	0.0969	0.2873	0.1939	0.0012
	Radial	-0.0902	0.1075	0.0087	0.8852

Таблица 5. Прогнозирование больших доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	0.1500	0.4013	0.2804	0.0006
	Poly	0.1473	0.3990	0.2779	0.0007
	Radial	0.1039	0.3613	0.2368	0.0039
Только sentiment в усеченной форме	Linear	-0.1431	0.1293	-0.0070	0.9325
	Poly	-0.1412	0.1312	-0.0051	0.9513
	Radial	-0.0926	0.1793	0.0442	0.5951
Только sentiment в значимой форме	Linear	-0.1480	0.1244	-0.0120	0.8855
	Poly	-0.1288	0.1437	0.0076	0.9272
	Radial	-0.1631	0.1091	-0.0275	0.7408
Финансы и усеченная форма sentiment	Linear	0.0009	0.2683	0.1371	0.0978
	Poly	-0.0289	0.2404	0.1077	0.1940
	Radial	-0.0334	0.2361	0.1032	0.2133
Финансы и значимая форма sentiment	Linear	-0.0482	0.2186	0.0868	0.2895
	Poly	-0.0289	0.2404	0.1077	0.1940
	Radial	-0.0334	0.2361	0.1032	0.2133

Таблица 6. Классификация по направлению движения акций с использованием методов машинного обучения и sentiment-анализа

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5320	0.6440	0.5880	+
	Random_forest	0.4651	0.6174	0.5412	-
	Bayes_Naive	0.4766	0.6625	0.5696	-
	Boosting	0.4806	0.6363	0.5585	-
	SVM-Linear	0.5384	0.6416	0.5900	+
	SVM-Poly	0.5418	0.6626	0.6022	+
	SVM-Radial	0.5144	0.6575	0.5859	+
Использование только новостной	Logit	0.4687	0.5772	0.5230	-
	Random_forest	0.4846	0.5794	0.5320	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
информации (текущая и первый лаг)	Bayes_Naive	0.4496	0.5899	0.5198	-
	Boosting	0.4848	0.6036	0.5442	-
	SVM-Linear	0.4982	0.5191	0.5086	-
	SVM-Poly	0.4797	0.5458	0.5128	-
	SVM-Radial	0.4447	0.6072	0.5260	-
Использование лагирующей новостной информации	Logit	0.4387	0.6029	0.5208	-
	Random_forest	0.4299	0.5244	0.4771	-
	Bayes_Naive	0.4028	0.5552	0.4790	-
	Boosting	0.4230	0.5391	0.4811	-
	SVM-Linear	0.4484	0.5873	0.5178	-
	SVM-Poly	0.4560	0.5939	0.5250	-
	SVM-Radial	0.4061	0.5785	0.4923	-
Оба канала информации (текущая и первый лаг)	Logit	0.5180	0.6641	0.5911	+
	Random_forest	0.5024	0.5719	0.5372	+
	Bayes_Naive	0.4975	0.6458	0.5717	-
	Boosting	0.5256	0.6219	0.5738	+
	SVM-Linear	0.5372	0.6652	0.6012	+
	SVM-Poly	0.5487	0.6619	0.6053	+
	SVM-Radial	0.5221	0.6579	0.5900	+
Оба канала информации (новости лагируемые)	Logit	0.5405	0.6538	0.5972	+
	Random_forest	0.5066	0.6043	0.5554	+
	Bayes_Naive	0.4218	0.6384	0.5301	-
	Boosting	0.5352	0.6225	0.5788	+
	SVM-Linear	0.5334	0.6590	0.5962	+
	SVM-Poly	0.5669	0.6335	0.6002	+
	SVM-Radial	0.4873	0.6378	0.5625	-

Таблица 7. Классификация по размаху движения с использованием ML и сентиментов

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4162	0.5851	0.5006	-
	Random_forest	0.4541	0.5774	0.5157	-
	Bayes_Naive	0.4662	0.6572	0.5617	-
	Boosting	0.4729	0.6260	0.5494	-
	SVM-Linear	0.4570	0.5645	0.5108	-
	SVM-Poly	0.4826	0.5816	0.5321	-
	SVM-Radial	0.5129	0.6449	0.5789	+
Использование только новостной информации (текущая и первый лаг)	Logit	0.4018	0.5564	0.4791	-
	Random_forest	0.4210	0.5721	0.4965	-
	Bayes_Naive	0.4593	0.5520	0.5056	-
	Boosting	0.4471	0.5399	0.4935	-
	SVM-Linear	0.4248	0.5457	0.4853	-
	SVM-Poly	0.4948	0.5042	0.4995	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование лагированной новостной информации	SVM-Radial	0.4288	0.5846	0.5067	-
	Logit	0.4327	0.5724	0.5026	-
	Random_forest	0.4560	0.5349	0.4954	-
	Bayes_Naive	0.4554	0.5823	0.5189	-
	Boosting	0.4308	0.6008	0.5158	-
	SVM-Linear	0.4303	0.5770	0.5036	-
	SVM-Poly	0.4538	0.5799	0.5168	-
	SVM-Radial	0.4582	0.5470	0.5026	-
Оба канала информации (текущая и первый лаг)	Logit	0.4245	0.5746	0.4996	-
	Random_forest	0.4743	0.6101	0.5422	-
	Bayes_Naive	0.5026	0.6063	0.5545	+
	Boosting	0.4627	0.6280	0.5454	-
	SVM-Linear	0.4443	0.5406	0.4924	-
	SVM-Poly	0.4419	0.5937	0.5178	-
	SVM-Radial	0.4903	0.6187	0.5545	-
Оба канала информации (новости лагированные)	Logit	0.4341	0.5835	0.5088	-
	Random_forest	0.4393	0.6452	0.5422	-
	Bayes_Naive	0.4429	0.6274	0.5351	-
	Boosting	0.4843	0.6452	0.5647	-
	SVM-Linear	0.4638	0.5658	0.5148	-
	SVM-Poly	0.4691	0.5707	0.5199	-
	SVM-Radial	0.4619	0.6348	0.5483	-

Таблица 8. ARIMA-модель для Киа с выделением топиков по методу LDA

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ar1	0.0215	0.0500	0.4290	0.6679
ar2	- 0.8886	0.0607	- 14.6420	0.0000
ar3	0.1976	0.0373	5.2937	0.0000
ar4	0.0061	0.0368	0.1658	0.8683
ma1	0.2337	0.0384	6.0822	0.0000
ma2	0.9187	0.0541	16.9745	0.0000
intercept	0.0009	0.0010	0.9313	0.3517
topik_4_2	- 0.0139	0.0073	- 1.8991	0.0576
topik_5_1	0.0170	0.0075	2.2574	0.0240
topik_8	- 0.0141	0.0066	- 2.1340	0.0328
topik_7_1	- 0.0116	0.0062	- 1.8727	0.0611
topik_5	0.0110	0.0077	1.4286	0.1531
topik_2_2	0.0075	0.0054	1.4077	0.1592
AIC	- 5 025			
BIC	- 4 957			
n.obs	985			

Статистически значимыми являются тематические группы, посвященные новостям об автомобилях данного бренда, об изменениях цен на автомобили Kia, безопасности данных автомобилей, а также динамике продаж и электромобилям.

Таблица 9. Прогнозы доходностей с помощью модели с использованием LDA

Источник: составлено автором

	int_a_arima_dir	int_b_arima_dir	cor_arima_dir	pv_arima_dir
kia	0.0023	0.2343	0.1199	0.0458

Таблица 10. Прогнозы дневных доходностей с помощью ML и LDA-словарей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Финансовая информация	Linear	-0.2009	0.0238	-0.0897	0.1211
	Poly	-0.1999	0.0248	-0.0887	0.1254
	Radial	-0.2255	-0.0020	-0.1152	0.0462
Только текущие словари	Linear	-0.1423	0.0840	-0.0296	0.6100
	Poly	-0.1186	0.1079	-0.0054	0.9254
	Radial	-0.1505	0.0757	-0.0379	0.5134
Только лаги словарей	Linear	-0.1571	0.0690	-0.0446	0.4414
	Poly	-0.1466	0.0796	-0.0339	0.5581
	Radial	-0.1091	0.1174	0.0042	0.9424
2 источника – текущие словари	Linear	-0.1635	0.0624	-0.0512	0.3770
	Poly	-0.1263	0.1001	-0.0133	0.8191
	Radial	-0.0932	0.1332	0.0202	0.7270
2 источника – лаги словарей	Linear	-0.1698	0.0559	-0.0577	0.3195
	Poly	-0.1493	0.0769	-0.0367	0.5266
	Radial	-0.1109	0.1155	0.0023	0.9680

Таблица 11. Классификация направления изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5320	0.6440	0.5880	+
	Random_forest	0.4651	0.6174	0.5412	-
	Bayes_Naive	0.4766	0.6625	0.5696	-
	Boosting	0.4806	0.6363	0.5585	-
	SVM-Linear	0.5384	0.6416	0.5900	+
	SVM-Poly	0.5418	0.6626	0.6022	+
	SVM-Radial	0.5144	0.6575	0.5859	+
	Logit	0.4596	0.5781	0.5188	-
	Random_forest	0.4563	0.5752	0.5158	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только новостной информации (текущая)	Bayes_Naive	0.4624	0.5915	0.5270	-
	Boosting	0.4506	0.5810	0.5158	-
	SVM-Linear	0.4636	0.5741	0.5188	-
	SVM-Poly	0.4608	0.5831	0.5219	-
	SVM-Radial	0.4695	0.5723	0.5209	-
Использование лагированной новостной информации	Logit	0.4514	0.5230	0.4872	-
	Random_forest	0.4416	0.5370	0.4893	-
	Bayes_Naive	0.4411	0.5721	0.5066	-
	Boosting	0.4361	0.5671	0.5016	-
	SVM-Linear	0.3999	0.5643	0.4821	-
	SVM-Poly	0.5049	0.5144	0.5097	+
	SVM-Radial	0.4532	0.5172	0.4852	-
Оба канала информации (текущая)	Logit	0.5296	0.6363	0.5830	+
	Random_forest	0.4882	0.6450	0.5666	-
	Bayes_Naive	0.4555	0.6433	0.5494	-
	Boosting	0.5046	0.6410	0.5728	+
	SVM-Linear	0.5259	0.6625	0.5942	+
	SVM-Poly	0.5253	0.6833	0.6043	+
	SVM-Radial	0.4791	0.6480	0.5635	-
Оба канала информации (новости лагированные)	Logit	0.4865	0.6304	0.5585	-
	Random_forest	0.4842	0.6162	0.5502	-
	Bayes_Naive	0.4418	0.5876	0.5147	-
	Boosting	0.5058	0.6416	0.5737	+
	SVM-Linear	0.4996	0.5948	0.5472	-
	SVM-Poly	0.4955	0.6133	0.5544	-
	SVM-Radial	0.4750	0.5870	0.5310	-

Таблица 12. Классификация магнитуды изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4162	0.5851	0.5006	-
	Random_forest	0.4541	0.5774	0.5157	-
	Bayes_Naive	0.4662	0.6572	0.5617	-
	Boosting	0.4729	0.6260	0.5494	-
	SVM-Linear	0.4570	0.5645	0.5108	-
	SVM-Poly	0.4826	0.5816	0.5321	-
	SVM-Radial	0.5129	0.6449	0.5789	+
Использование только новостной информации (текущая)	Logit	0.4230	0.5576	0.4903	-
	Random_forest	0.4237	0.5041	0.4639	-
	Bayes_Naive	0.4304	0.5665	0.4985	-
	Boosting	0.4143	0.5481	0.4812	-
	SVM-Linear	0.4205	0.5520	0.4863	-
	SVM-Poly	0.4815	0.5317	0.5066	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование лагированной новостной информации	SVM-Radial	0.4180	0.5423	0.4802	-
	Logit	0.4598	0.5980	0.5289	-
	Random_forest	0.4465	0.6256	0.5361	-
	Bayes_Naive	0.4407	0.6213	0.5310	-
	Boosting	0.4291	0.5983	0.5137	-
	SVM-Linear	0.4336	0.6161	0.5249	-
	SVM-Poly	0.4363	0.6317	0.5340	-
	SVM-Radial	0.4667	0.5994	0.5330	-
Оба канала информации (текущая)	Logit	0.4261	0.5852	0.5057	-
	Random_forest	0.4655	0.5803	0.5229	-
	Bayes_Naive	0.4871	0.6013	0.5442	-
	Boosting	0.4593	0.6519	0.5556	-
	SVM-Linear	0.4309	0.5355	0.4832	-
	SVM-Poly	0.4376	0.6144	0.5260	-
	SVM-Radial	0.4655	0.5987	0.5321	-
Оба канала информации (новости лагированные)	Logit	0.4774	0.5907	0.5341	-
	Random_forest	0.4604	0.6670	0.5637	-
	Bayes_Naive	0.4621	0.6080	0.5350	-
	Boosting	0.4644	0.6506	0.5575	-
	SVM-Linear	0.4811	0.6114	0.5463	-
	SVM-Poly	0.4605	0.6138	0.5372	-
	SVM-Radial	0.4602	0.6120	0.5361	-

Таблица 13. Классификация направления изменения доходности по важным для компании
Киа словам (TF-IDF)

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5320	0.6440	0.5880	+
	Random_forest	0.4651	0.6174	0.5412	-
	Bayes_Naive	0.4766	0.6625	0.5696	-
	Boosting	0.4806	0.6363	0.5585	-
	SVM-Linear	0.5384	0.6416	0.5900	+
	SVM-Poly	0.5418	0.6626	0.6022	+
	SVM-Radial	0.5144	0.6575	0.5859	+
Использование только новостной информации (текущая)	Logit	0.4759	0.5415	0.5087	-
	Random_forest	0.4714	0.5582	0.5148	-
	Bayes_Naive	0.4863	0.5208	0.5036	-
	Boosting	0.4795	0.5582	0.5188	-
	SVM-Linear	0.4884	0.5596	0.5240	-
	SVM-Poly	0.4940	0.5417	0.5178	-
	SVM-Radial	0.4646	0.5671	0.5159	-
	Logit	0.4647	0.5578	0.5112	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование лагированной новостной информации	Random_forest	0.4979	0.5287	0.5133	-
	Bayes_Naive	0.4971	0.5233	0.5102	-
	Boosting	0.4803	0.5625	0.5214	-
	SVM-Linear	0.4853	0.5474	0.5163	-
	SVM-Poly	0.4913	0.5413	0.5163	-
	SVM-Radial	0.4657	0.5650	0.5153	-
Оба канала информации (текущая)	Logit	0.5241	0.6519	0.5880	+
	Random_forest	0.4790	0.6055	0.5423	-
	Bayes_Naive	0.4881	0.5230	0.5056	-
	Boosting	0.5074	0.6400	0.5737	+
	SVM-Linear	0.4889	0.5529	0.5209	-
	SVM-Poly	0.4941	0.5396	0.5168	-
	SVM-Radial	0.4825	0.6589	0.5707	-
Оба канала информации (новости лагированные)	Logit	0.4961	0.6752	0.5857	-
	Random_forest	0.4729	0.5985	0.5357	-
	Bayes_Naive	0.4932	0.5293	0.5112	-
	Boosting	0.4727	0.6803	0.5765	-
	SVM-Linear	0.4775	0.5532	0.5153	-
	SVM-Poly	0.4913	0.5413	0.5163	-
	SVM-Radial	0.4438	0.6644	0.5541	-

Таблица 14. Классификация масштаба доходности по важным для компании словам

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4162	0.5851	0.5006	-
	Random_forest	0.4541	0.5774	0.5157	-
	Bayes_Naive	0.4662	0.6572	0.5617	-
	Boosting	0.4729	0.6260	0.5494	-
	SVM-Linear	0.4570	0.5645	0.5108	-
	SVM-Poly	0.4826	0.5816	0.5321	-
	SVM-Radial	0.5129	0.6449	0.5789	+
Использование только новостной информации (текущая)	Logit	0.4644	0.5428	0.5036	-
	Random_forest	0.4670	0.5525	0.5097	-
	Bayes_Naive	0.4758	0.5334	0.5046	-
	Boosting	0.4399	0.5266	0.4833	-
	SVM-Linear	0.4648	0.5301	0.4975	-
	SVM-Poly	0.4945	0.5025	0.4985	-
	SVM-Radial	0.4416	0.5719	0.5068	-
Использование лагированной новостной информации	Logit	0.4747	0.5395	0.5071	-
	Random_forest	0.4740	0.5484	0.5112	-
	Bayes_Naive	0.4713	0.5308	0.5011	-
	Boosting	0.4460	0.5194	0.4827	-
	SVM-Linear	0.4685	0.5417	0.5051	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
	SVM-Poly	0.4723	0.5298	0.5010	-
	SVM-Radial	0.4856	0.5409	0.5132	-
Оба канала информации (текущая)	Logit	0.4286	0.5683	0.4985	-
	Random_forest	0.4516	0.6005	0.5260	-
	Bayes_Naive	0.4631	0.5564	0.5097	-
	Boosting	0.4739	0.6678	0.5708	-
	SVM-Linear	0.4648	0.5301	0.4975	-
	SVM-Poly	0.4945	0.5025	0.4985	-
	SVM-Radial	0.5112	0.5895	0.5504	+
Оба канала информации (новости ланированные)	Logit	0.4417	0.5542	0.4980	-
	Random_forest	0.4442	0.5926	0.5184	-
	Bayes_Naive	0.4693	0.5369	0.5031	-
	Boosting	0.4901	0.6261	0.5581	-
	SVM-Linear	0.4691	0.5452	0.5071	-
	SVM-Poly	0.4723	0.5298	0.5010	-
	SVM-Radial	0.4811	0.6453	0.5632	-

Результаты проведенного анализа по компании General Motors

Таблица 1. ARIMA-модель для модели с применением сентимент-анализа

Источник: составлено автором

	estimate	Sd	t_stat	pi_val
ar1	− 0.2994	0.4893	− 0.6118	0.5407
ar2	0.3269	0.5971	0.5474	0.5841
ar3	− 0.1003	0.4519	− 0.2219	0.8244
ma1	0.6077	0.4882	1.2448	0.2132
ma2	− 0.1918	0.7175	− 0.2673	0.7893
ma3	− 0.0185	0.5391	− 0.0342	0.9727
ma4	0.0346	0.1501	0.2304	0.8178
ma5	0.0854	0.0371	2.3009	0.0214
intercept	0.0002	0.0012	0.1395	0.8891
sent	0.0005	0.0004	1.2445	0.2133
sent_1	− 0.0000	0.0004	− 0.0246	0.9804
sent_4	− 0.0003	0.0004	− 0.6422	0.5207
AIC	− 4,293			
BIC	− 4,229			
n.obs	986			

Таблица 2. Прогнозирование доходностей акций компании с помощью ARIMA-модели на всей выборке

Источник: составлено автором

	int_a_arima	int_b_arima	cor_arima	pv_arima
genmt	−0.0285	0.1610	0.0669	0.2482

Таблица 3. Прогнозирование больших по модулю доходностей с использованием сентимент-анализа

Источник: Составлено автором

	int_a_lm	int_b_lm	cor_lm	pv_lm
genmt	−0.0084	0.2562	0.1261	0.1228

Таблица 4. Прогнозирование доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	0.1446	0.3243	0.2365	0.0000
	Poly	0.1446	0.3243	0.2365	0.0000
	Radial	0.0896	0.2736	0.1832	0.0014
	Linear	−0.0270	0.1625	0.0684	0.2377

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только sentiment в усеченной форме	Poly	-0.0292	0.1603	0.0661	0.2535
	Radial	-0.0232	0.1662	0.0722	0.2127
Только sentiment в значимой форме	Linear	-0.0285	0.1610	0.0669	0.2482
	Poly	0.0401	0.2269	0.1347	0.0196
	Radial	-0.0701	0.1201	0.0252	0.6632
Финансы + усеченная форма sentiment	Linear	0.1560	0.3347	0.2475	0.0000
	Poly	0.1570	0.3357	0.2484	0.0000
	Radial	-0.0045	0.1842	0.0907	0.1170
Финансы + значимая форма sentiment	Linear	0.1432	0.3230	0.2351	0.0000
	Poly	0.1461	0.3257	0.2379	0.0000
	Radial	0.1307	0.3117	0.2231	0.0001

Таблица 5. Прогнозирование больших доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	0.1812	0.4249	0.3081	0.0001
	Poly	0.1842	0.4274	0.3109	0.0001
	Radial	0.1663	0.4122	0.2941	0.0002
Только sentiment в усеченной форме	Linear	-0.0439	0.2227	0.0911	0.2661
	Poly	-0.0387	0.2276	0.0961	0.2403
	Radial	-0.0556	0.2115	0.0794	0.3324
Только sentiment в значимой форме	Linear	-0.0084	0.2562	0.1261	0.1228
	Poly	-0.0084	0.2562	0.1261	0.1228
	Radial	-0.0494	0.2175	0.0856	0.2960
Финансы и усеченная форма sentiment	Linear	0.1933	0.4351	0.3193	0.0001
	Poly	0.2183	0.4561	0.3427	0.0000
	Radial	0.1785	0.4226	0.3056	0.0001
Финансы и значимая форма sentiment	Linear	0.2170	0.4549	0.3414	0.0000
	Poly	0.2188	0.4565	0.3431	0.0000
	Radial	0.2504	0.4826	0.3723	0.0000

Таблица 6. Классификация по направлению движения акций с использованием методов машинного обучения и sentiment-анализа

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5106	0.6275	0.5691	+
	Random_forest	0.4426	0.6539	0.5483	-
	Bayes_Naive	0.4673	0.6309	0.5491	-
	Boosting	0.5089	0.5955	0.5522	+
	SVM-Linear	0.5054	0.6208	0.5631	+
	SVM-Poly	0.4997	0.6265	0.5631	-
	SVM-Radial	0.4949	0.6174	0.5562	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только новостной информации (текущая и первый лаг)	Logit	0.4609	0.5501	0.5055	-
	Random_forest	0.4334	0.4919	0.4627	-
	Bayes_Naive	0.4528	0.5661	0.5094	-
	Boosting	0.4265	0.5486	0.4875	-
	SVM-Linear	0.4949	0.5241	0.5095	-
	SVM-Poly	0.4904	0.5405	0.5154	-
	SVM-Radial	0.4021	0.5470	0.4746	-
Использование лагированной новостной информации	Logit	0.4598	0.6109	0.5354	-
	Random_forest	0.3816	0.6495	0.5155	-
	Bayes_Naive	0.4298	0.6191	0.5245	-
	Boosting	0.4514	0.5497	0.5006	-
	SVM-Linear	0.4853	0.5396	0.5125	-
	SVM-Poly	0.4895	0.5612	0.5254	-
	SVM-Radial	0.4401	0.5670	0.5035	-
Оба канала информации (текущая и первый лаг)	Logit	0.4893	0.6369	0.5631	-
	Random_forest	0.4243	0.6305	0.5274	-
	Bayes_Naive	0.4660	0.5807	0.5233	-
	Boosting	0.4887	0.6077	0.5482	-
	SVM-Linear	0.5017	0.6067	0.5542	+
	SVM-Poly	0.5169	0.6054	0.5611	+
	SVM-Radial	0.4867	0.6196	0.5532	-
Оба канала информации (новости лагированные)	Logit	0.4807	0.6336	0.5572	-
	Random_forest	0.4452	0.6634	0.5543	-
	Bayes_Naive	0.4911	0.5876	0.5393	-
	Boosting	0.4882	0.6241	0.5561	-
	SVM-Linear	0.4715	0.6270	0.5492	-
	SVM-Poly	0.5003	0.6280	0.5641	+
	SVM-Radial	0.4914	0.5873	0.5394	-

Таблица 7. Классификация по размаху движения с использованием ML и сентиментов

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4398	0.6089	0.5244	-
	Random_forest	0.4798	0.5847	0.5323	-
	Bayes_Naive	0.5053	0.6311	0.5682	+
	Boosting	0.4954	0.6350	0.5652	-
	SVM-Linear	0.4709	0.5977	0.5343	-
	SVM-Poly	0.4690	0.6155	0.5423	-
	SVM-Radial	0.4948	0.6693	0.5821	-
Использование только новостной информации	Logit	0.4757	0.5849	0.5303	-
	Random_forest	0.4523	0.5647	0.5085	-
	Bayes_Naive	0.4692	0.5696	0.5194	-
	Boosting	0.4416	0.5455	0.4936	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
(текущая и первый лаг)	SVM-Linear	0.4775	0.5573	0.5174	-
	SVM-Poly	0.4740	0.5609	0.5174	-
	SVM-Radial	0.4426	0.5643	0.5035	-
Использование лагированной новостной информации	Logit	0.4261	0.5550	0.4905	-
	Random_forest	0.4564	0.5865	0.5214	-
	Bayes_Naive	0.4520	0.5570	0.5045	-
	Boosting	0.4539	0.5889	0.5214	-
	SVM-Linear	0.4702	0.5268	0.4985	-
	SVM-Poly	0.4717	0.5372	0.5045	-
	SVM-Radial	0.4745	0.5984	0.5365	-
Оба канала информации (текущая и первый лаг)	Logit	0.4806	0.6318	0.5562	-
	Random_forest	0.4781	0.6342	0.5561	-
	Bayes_Naive	0.4964	0.6181	0.5572	-
	Boosting	0.5021	0.6500	0.5761	+
	SVM-Linear	0.4850	0.5737	0.5294	-
	SVM-Poly	0.4727	0.6019	0.5373	-
	SVM-Radial	0.5035	0.6627	0.5831	+
Оба канала информации (новости ланированные)	Logit	0.4330	0.6019	0.5174	-
	Random_forest	0.4959	0.6106	0.5532	-
	Bayes_Naive	0.4820	0.6226	0.5523	-
	Boosting	0.4837	0.6448	0.5643	-
	SVM-Linear	0.4663	0.5844	0.5254	-
	SVM-Poly	0.4664	0.6101	0.5383	-
	SVM-Radial	0.4836	0.6587	0.5711	-

Таблица 8. ARIMA-модель для Дженерал Моторс с выделением топиков по методу LDA

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ar1	0.1998	0.0309	6.4707	0.0000
intercept	- 0.0010	0.0010	- 0.9830	0.3256
topik_7	0.0155	0.0082	1.8968	0.0579
AIC	- 4 724			
BIC	- 4 700			
n.obs	1 007			

Статистически значимой является только 1 тематическая группа, посвященная продажам компании.

Таблица 9. Прогнозы доходностей с помощью модели с использованием LDA

Источник: составлено автором

	int_a_arima_dir	int_b_arima_dir	cor_arima_dir	pv_arima_dir
genmt	- 0.0384	0.1868	0.0751	0.1944

Таблица 10. Прогнозы дневных доходностей с помощью ML и LDA-словарей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Финансовая информация	Linear	0.1452	0.3572	0.2543	0.0000
	Poly	0.1423	0.3547	0.2515	0.0000
	Radial	0.1150	0.3301	0.2253	0.0001
Только текущие словари	Linear	-0.0487	0.1768	0.0649	0.2626
	Poly	-0.0328	0.1922	0.0807	0.1632
	Radial	-0.1114	0.1151	0.0019	0.9737
Только лаги словарей	Linear	-0.1405	0.0858	-0.0277	0.6330
	Poly	-0.1944	0.0306	-0.0829	0.1518
	Radial	-0.1536	0.0725	-0.0410	0.4788
2 источника – текущие словари	Linear	0.1149	0.3300	0.2252	0.0001
	Poly	0.1236	0.3378	0.2335	0.0000
	Radial	0.0753	0.2940	0.1870	0.0011
2 источника – лаги словарей	Linear	0.0532	0.2735	0.1654	0.0041
	Poly	0.0021	0.2256	0.1154	0.0459
	Radial	-0.0142	0.2101	0.0992	0.0863

Таблица 11. Классификация направления изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5106	0.6275	0.5691	+
	Random_forest	0.4426	0.6539	0.5483	-
	Bayes_Naive	0.4673	0.6309	0.5491	-
	Boosting	0.5089	0.5955	0.5522	+
	SVM-Linear	0.5054	0.6208	0.5631	+
	SVM-Poly	0.4997	0.6265	0.5631	-
	SVM-Radial	0.4949	0.6174	0.5562	-
Использование только новостной информации (текущая)	Logit	0.4406	0.6064	0.5235	-
	Random_forest	0.4103	0.5947	0.5025	-
	Bayes_Naive	0.4340	0.6009	0.5174	-
	Boosting	0.4240	0.5930	0.5085	-
	SVM-Linear	0.4433	0.6057	0.5245	-
	SVM-Poly	0.4482	0.5987	0.5235	-
	SVM-Radial	0.4420	0.6010	0.5215	-
	Logit	0.4385	0.5744	0.5065	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование лагированной новостной информации	Random_forest	0.4194	0.6053	0.5123	-
	Bayes_Naive	0.4437	0.5911	0.5174	-
	Boosting	0.4028	0.5901	0.4965	-
	SVM-Linear	0.4607	0.5362	0.4985	-
	SVM-Poly	0.4527	0.5883	0.5205	-
	SVM-Radial	0.4018	0.5993	0.5006	-
Оба канала информации (текущая)	Logit	0.4811	0.6272	0.5542	-
	Random_forest	0.4773	0.6174	0.5473	-
	Bayes_Naive	0.4177	0.5993	0.5085	-
	Boosting	0.5006	0.6099	0.5553	+
	SVM-Linear	0.4881	0.6225	0.5553	-
	SVM-Poly	0.4657	0.6469	0.5563	-
	SVM-Radial	0.4790	0.6295	0.5543	-
Оба канала информации (новости лагированные)	Logit	0.4987	0.6059	0.5523	-
	Random_forest	0.4371	0.6295	0.5333	-
	Bayes_Naive	0.4625	0.6061	0.5343	-
	Boosting	0.4673	0.6787	0.5730	-
	SVM-Linear	0.4317	0.6370	0.5344	-
	SVM-Poly	0.4894	0.5952	0.5423	-
	SVM-Radial	0.4749	0.5960	0.5354	-

Таблица 12. Классификация магнитуды изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5106	0.6275	0.5691	+
	Random_forest	0.4426	0.6539	0.5483	-
	Bayes_Naive	0.4673	0.6309	0.5491	-
	Boosting	0.5089	0.5955	0.5522	+
	SVM-Linear	0.5054	0.6208	0.5631	+
	SVM-Poly	0.4997	0.6265	0.5631	-
	SVM-Radial	0.4949	0.6174	0.5562	-
Использование только новостной информации (текущая)	Logit	0.4609	0.5501	0.5055	-
	Random_forest	0.4334	0.4919	0.4627	-
	Bayes_Naive	0.4528	0.5661	0.5094	-
	Boosting	0.4265	0.5486	0.4875	-
	SVM-Linear	0.4949	0.5241	0.5095	-
	SVM-Poly	0.4904	0.5405	0.5154	-
	SVM-Radial	0.4021	0.5470	0.4746	-
Использование лагированной новостной информации	Logit	0.4598	0.6109	0.5354	-
	Random_forest	0.3816	0.6495	0.5155	-
	Bayes_Naive	0.4298	0.6191	0.5245	-
	Boosting	0.4514	0.5497	0.5006	-
	SVM-Linear	0.4853	0.5396	0.5125	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (текущая)	SVM-Poly	0.4895	0.5612	0.5254	-
	SVM-Radial	0.4401	0.5670	0.5035	-
	Logit	0.4893	0.6369	0.5631	-
	Random_forest	0.4243	0.6305	0.5274	-
	Bayes_Naive	0.4660	0.5807	0.5233	-
	Boosting	0.4887	0.6077	0.5482	-
	SVM-Linear	0.5017	0.6067	0.5542	+
Оба канала информации (новости лагированные)	SVM-Poly	0.5169	0.6054	0.5611	+
	SVM-Radial	0.4867	0.6196	0.5532	-
	Logit	0.4807	0.6336	0.5572	-
	Random_forest	0.4452	0.6634	0.5543	-
	Bayes_Naive	0.4911	0.5876	0.5393	-
	Boosting	0.4882	0.6241	0.5561	-
	SVM-Linear	0.4715	0.6270	0.5492	-
	SVM-Poly	0.5003	0.6280	0.5641	+
	SVM-Radial	0.4914	0.5873	0.5394	-

Таблица 13. Классификация направления изменения доходности по важным для компании
Дженерал Моторс словам (TF-IDF)

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5106	0.6275	0.5691	+
	Random_forest	0.4426	0.6539	0.5483	-
	Bayes_Naive	0.4673	0.6309	0.5491	-
	Boosting	0.5089	0.5955	0.5522	+
	SVM-Linear	0.5054	0.6208	0.5631	+
	SVM-Poly	0.4997	0.6265	0.5631	-
	SVM-Radial	0.4949	0.6174	0.5562	-
Использование только новостной информации (текущая)	Logit	0.4500	0.5649	0.5075	-
	Random_forest	0.5060	0.5169	0.5114	+
	Bayes_Naive	0.4786	0.5342	0.5064	-
	Boosting	0.4720	0.5430	0.5075	-
	SVM-Linear	0.4688	0.5342	0.5015	-
	SVM-Poly	0.5060	0.5169	0.5114	+
	SVM-Radial	0.4708	0.5520	0.5114	-
Использование лагированной новостной информации	Logit	0.4889	0.5430	0.5160	-
	Random_forest	0.5035	0.5184	0.5109	+
	Bayes_Naive	0.4900	0.5319	0.5109	-
	Boosting	0.4851	0.5310	0.5080	-
	SVM-Linear	0.4864	0.5414	0.5139	-
	SVM-Poly	0.5078	0.5161	0.5119	+
	SVM-Radial	0.4646	0.5632	0.5139	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (текущая)	Logit	0.4803	0.6540	0.5671	-
	Random_forest	0.4467	0.6200	0.5333	-
	Bayes_Naive	0.4786	0.5342	0.5064	-
	Boosting	0.4930	0.6054	0.5492	-
	SVM-Linear	0.4628	0.5421	0.5025	-
	SVM-Poly	0.5060	0.5169	0.5114	+
	SVM-Radial	0.5034	0.6287	0.5661	+
Оба канала информации (новости лагированные)	Logit	0.5043	0.6330	0.5686	+
	Random_forest	0.4965	0.6170	0.5567	-
	Bayes_Naive	0.4900	0.5319	0.5109	-
	Boosting	0.4904	0.6210	0.5557	-
	SVM-Linear	0.4808	0.5510	0.5159	-
	SVM-Poly	0.5078	0.5161	0.5119	+
	SVM-Radial	0.4900	0.6531	0.5716	-

Таблица 14. Классификация масштаба доходности по важным для компании словам

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4398	0.6089	0.5244	-
	Random_forest	0.4798	0.5847	0.5323	-
	Bayes_Naive	0.5053	0.6311	0.5682	+
	Boosting	0.4954	0.6350	0.5652	-
	SVM-Linear	0.4709	0.5977	0.5343	-
	SVM-Poly	0.4690	0.6155	0.5423	-
	SVM-Radial	0.4948	0.6693	0.5821	-
Использование только новостной информации (текущая)	Logit	0.4853	0.5356	0.5104	-
	Random_forest	0.4895	0.5354	0.5125	-
	Bayes_Naive	0.4903	0.5365	0.5134	-
	Boosting	0.4823	0.5506	0.5164	-
	SVM-Linear	0.4888	0.5301	0.5095	-
	SVM-Poly	0.4921	0.5388	0.5154	-
	SVM-Radial	0.4859	0.5390	0.5125	-
Использование лагированной новостной информации	Logit	0.4881	0.5397	0.5139	-
	Random_forest	0.4803	0.5356	0.5080	-
	Bayes_Naive	0.4792	0.5366	0.5079	-
	Boosting	0.5011	0.5308	0.5159	+
	SVM-Linear	0.4890	0.5368	0.5129	-
	SVM-Poly	0.4919	0.5400	0.5159	-
	SVM-Radial	0.4828	0.5411	0.5119	-
Оба канала информации (текущая)	Logit	0.4583	0.6064	0.5323	-
	Random_forest	0.4789	0.5797	0.5293	-
	Bayes_Naive	0.4903	0.5365	0.5134	-
	Boosting	0.4849	0.6355	0.5602	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
	SVM-Linear	0.4884	0.5325	0.5105	-
	SVM-Poly	0.4971	0.5338	0.5154	-
	SVM-Radial	0.4730	0.6673	0.5702	-
Оба канала информации (новости ланированные)	Logit	0.4402	0.6215	0.5308	-
	Random_forest	0.4784	0.5892	0.5338	-
	Bayes_Naive	0.4835	0.5423	0.5129	-
	Boosting	0.5302	0.6412	0.5857	+
	SVM-Linear	0.4910	0.5349	0.5129	-
	SVM-Poly	0.4919	0.5400	0.5159	-
	SVM-Radial	0.5069	0.6324	0.5697	+

Результаты проведенного анализа по компании Toyota



Таблица 1. ARIMA-модель для модели с применением сентимент-анализа

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ar1	0.9118	0.0354	25.7777	0.0000
ar2	− 0.9175	0.0512	− 17.9244	0.0000
ma1	− 0.8553	0.0410	− 20.8392	0.0000
ma2	0.9022	0.0600	15.0244	0.0000
intercept	0.0002	0.0004	0.4856	0.6272
sent	0.0001	0.0001	1.3089	0.1906
AIC	− 5 870			
BIC	− 5 836			
n.obs	1 007			

Таблица 2. Прогнозирование больших по модулю доходностей с использованием сентимент-анализа

Источник: Составлено автором

	int_a_lm	int_b_lm	cor_lm	pv_lm
toyota	− 0.0871	0.1811	0.0479	0.5595

Таблица 3. Прогнозирование доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	0.0594	0.2452	0.1537	0.0077
	Poly	0.0189	0.2068	0.1139	0.0487
	Radial	-0.0765	0.1138	0.0188	0.7454
Только сентимент в усеченной форме	Linear	-0.1642	0.0251	-0.0702	0.2255
	Poly	-0.1612	0.0283	-0.0670	0.2470
	Radial	-0.1416	0.0483	-0.0471	0.4164
Только сентимент в значимой форме	Linear	-0.0366	0.1530	0.0587	0.3105
	Poly	-0.0488	0.1411	0.0466	0.4213
	Radial	-0.2091	-0.0213	-0.1163	0.0442
Финансы + усеченная форма сентимента	Linear	-0.0448	0.1450	0.0506	0.3828
	Poly	-0.0824	0.1079	0.0128	0.8246
	Radial	-0.0413	0.1484	0.0541	0.3508
Финансы + значимая форма сентимента	Linear	0.0531	0.2393	0.1475	0.0105
	Poly	-0.0001	0.1885	0.0950	0.1004
	Radial	-0.0720	0.1182	0.0233	0.6881

Таблица 4. Прогнозирование больших доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	0.0602	0.3191	0.1930	0.0176
	Poly	0.0654	0.3238	0.1980	0.0148
	Radial	0.0366	0.2977	0.1701	0.0368
Только сентимент в усеченной форме	Linear	-0.1258	0.1429	0.0087	0.9156
	Poly	-0.2175	0.0493	-0.0856	0.2959
	Radial	-0.2208	0.0459	-0.0890	0.2770
Только сентимент в значимой форме	Linear	-0.0841	0.1840	0.0509	0.5351
	Poly	-0.1666	0.1019	-0.0329	0.6880
	Radial	-0.2278	0.0386	-0.0963	0.2394
Финансы и усеченная форма сентимента	Linear	0.0305	0.2921	0.1642	0.0439
	Poly	0.0202	0.2827	0.1541	0.0588
	Radial	0.1002	0.3548	0.2314	0.0042
Финансы и значимая форма сентимента	Linear	0.0522	0.3119	0.1853	0.0227
	Poly	-0.0262	0.2395	0.1086	0.1846
	Radial	-0.0085	0.2561	0.1261	0.1230

Таблица 5. Классификация по направлению движения акций с использованием методов машинного обучения и сентимент-анализа

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4740	0.6045	0.5393	-
	Random_forest	0.4542	0.5568	0.5055	-
	Bayes_Naive	0.4457	0.6090	0.5273	-
	Boosting	0.4277	0.6171	0.5224	-
	SVM-Linear	0.4897	0.5450	0.5174	-
	SVM-Poly	0.5183	0.5265	0.5224	+
	SVM-Radial	0.4236	0.5893	0.5065	-
Использование только новостной информации (текущая и первый лаг)	Logit	0.4966	0.5601	0.5283	-
	Random_forest	0.4393	0.5655	0.5024	-
	Bayes_Naive	0.4695	0.5414	0.5055	-
	Boosting	0.4634	0.5535	0.5084	-
	SVM-Linear	0.5183	0.5265	0.5224	+
	SVM-Poly	0.5183	0.5265	0.5224	+
	SVM-Radial	0.4784	0.5404	0.5094	-
Использование лагированной новостной информации	Logit	0.4637	0.5531	0.5084	-
	Random_forest	0.4200	0.5374	0.4787	-
	Bayes_Naive	0.4039	0.5434	0.4736	-
	Boosting	0.4092	0.5680	0.4886	-
	SVM-Linear	0.5183	0.5265	0.5224	+
	SVM-Poly	0.5183	0.5265	0.5224	+

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (текущая и первый лаг)	SVM-Radial	0.4676	0.5474	0.5075	-
	Logit	0.4828	0.5917	0.5372	-
	Random_forest	0.4120	0.5730	0.4925	-
	Bayes_Naive	0.4335	0.5875	0.5105	-
	Boosting	0.4037	0.5991	0.5014	-
	SVM-Linear	0.4656	0.5832	0.5244	-
	SVM-Poly	0.5183	0.5265	0.5224	+
	SVM-Radial	0.4482	0.5826	0.5154	-
Оба канала информации (новости лагированные)	Logit	0.4505	0.6021	0.5263	-
	Random_forest	0.4345	0.5365	0.4855	-
	Bayes_Naive	0.4397	0.5653	0.5025	-
	Boosting	0.3954	0.5955	0.4954	-
	SVM-Linear	0.4676	0.5552	0.5114	-
	SVM-Poly	0.5183	0.5265	0.5224	+
	SVM-Radial	0.4274	0.6073	0.5173	-

Таблица 6. Классификация по размаху движения с использованием ML и сентиментов

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4522	0.5488	0.5005	-
	Random_forest	0.4752	0.5695	0.5223	-
	Bayes_Naive	0.4726	0.6336	0.5531	-
	Boosting	0.4899	0.6026	0.5462	-
	SVM-Linear	0.4756	0.5770	0.5263	-
	SVM-Poly	0.4795	0.5870	0.5333	-
	SVM-Radial	0.4929	0.6096	0.5513	-
Использование только новостной информации (текущая и первый лаг)	Logit	0.4312	0.5201	0.4756	-
	Random_forest	0.4638	0.5968	0.5303	-
	Bayes_Naive	0.4447	0.5522	0.4985	-
	Boosting	0.4291	0.5540	0.4915	-
	SVM-Linear	0.4365	0.5227	0.4796	-
	SVM-Poly	0.4790	0.5121	0.4955	-
	SVM-Radial	0.4733	0.5794	0.5263	-
Использование лагированной новостной информации	Logit	0.4466	0.5724	0.5095	-
	Random_forest	0.4554	0.6293	0.5424	-
	Bayes_Naive	0.4551	0.5618	0.5085	-
	Boosting	0.4846	0.5881	0.5363	-
	SVM-Linear	0.4262	0.5709	0.4985	-
	SVM-Poly	0.4367	0.5884	0.5126	-
	SVM-Radial	0.4439	0.5910	0.5175	-
Оба канала информации	Logit	0.4189	0.5562	0.4875	-
	Random_forest	0.4478	0.6286	0.5382	-
	Bayes_Naive	0.4641	0.6283	0.5462	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
(текущая и первый лаг)	Boosting	0.4912	0.6370	0.5641	-
	SVM-Linear	0.4326	0.5802	0.5064	-
	SVM-Poly	0.4739	0.5927	0.5333	-
	SVM-Radial	0.4549	0.6096	0.5322	-
Оба канала информации (новости ланированные)	Logit	0.4570	0.5939	0.5254	-
	Random_forest	0.4326	0.6122	0.5224	-
	Bayes_Naive	0.4905	0.6060	0.5482	-
	Boosting	0.5066	0.6295	0.5681	+
	SVM-Linear	0.4344	0.6145	0.5245	-
	SVM-Poly	0.4880	0.5846	0.5363	-
	SVM-Radial	0.4661	0.6186	0.5423	-

Таблица 7. ARIMA-модель для Тойота с выделением топиков по методу LDA

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ar1	0.9065	0.0441	20.5458	0.0000
ar2	- 0.9032	0.0602	- 15.0016	0.0000
ma1	- 0.8455	0.0526	- 16.0629	0.0000
ma2	0.8867	0.0662	13.4021	0.0000
intercept	0.0015	0.0007	2.1888	0.0286
topik_1	- 0.0122	0.0042	- 2.9240	0.0035
topik_4_2	0.0066	0.0033	2.0146	0.0440
topik_9_2	- 0.0061	0.0029	- 2.1161	0.0343
topik_2_3	- 0.0115	0.0046	- 2.4796	0.0132
topik_1_3	0.0072	0.0042	1.7181	0.0858
topik_8	0.0078	0.0041	1.8972	0.0578
topik_3_2	- 0.0088	0.0041	- 2.1408	0.0323
topik_8_3	- 0.0080	0.0042	- 1.9206	0.0548
AIC	- 5 890			
BIC	- 5 821			
n.obs	1 007			

Статистически значимыми тематическими группами новостей являются: материалы, посвященные компании Тойота как японскому автопроизводителю, также производству электромобилей под брендом Тойота, их модельному ряду в целом, внедорожникам в частности, а также китайскому авто-рынку.

Таблица 8. Прогнозы доходностей с помощью модели с использованием LDA

Источник: составлено автором

	int_a_arima_dir	int_b_arima_dir	cor_arima_dir	pv_arima_dir
toyota	- 0.0021	0.2216	0.1112	0.0544

Таблица 9. Прогнозы дневных доходностей с помощью ML и LDA-словарей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Финансовая информация	Linear	-0.0413	0.1840	0.0723	0.2119
	Poly	-0.0677	0.1583	0.0459	0.4282
	Radial	-0.0319	0.1931	0.0816	0.1586
Только текущие словари	Linear	-0.0066	0.2174	0.1067	0.0648
	Poly	-0.0757	0.1505	0.0379	0.5136
	Radial	-0.0742	0.1519	0.0394	0.4970
Только лаги словарей	Linear	-0.1011	0.1254	0.0123	0.8318
	Poly	-0.1020	0.1244	0.0114	0.8447
	Radial	-0.0440	0.1814	0.0696	0.2293
2 источника – текущие словари	Linear	-0.0068	0.2172	0.1066	0.0653
	Poly	-0.0274	0.1974	0.0861	0.1367
	Radial	-0.0574	0.1683	0.0562	0.3323
2 источника – лаги словарей	Linear	-0.0869	0.1394	0.0266	0.6465
	Poly	-0.0865	0.1398	0.0270	0.6416
	Radial	-0.0306	0.1943	0.0829	0.1519

Таблица 10. Классификация направления изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4740	0.6045	0.5393	-
	Random_forest	0.4542	0.5568	0.5055	-
	Bayes_Naive	0.4457	0.6090	0.5273	-
	Boosting	0.4277	0.6171	0.5224	-
	SVM-Linear	0.4897	0.5450	0.5174	-
	SVM-Poly	0.5183	0.5265	0.5224	+
	SVM-Radial	0.4236	0.5893	0.5065	-
Использование только новостной информации (текущая)	Logit	0.4418	0.6132	0.5275	-
	Random_forest	0.4052	0.6462	0.5257	-
	Bayes_Naive	0.4372	0.6378	0.5375	-
	Boosting	0.4399	0.6173	0.5286	-
	SVM-Linear	0.4418	0.6292	0.5355	-
	SVM-Poly	0.4492	0.6217	0.5355	-
	SVM-Radial	0.4519	0.6130	0.5325	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование лагированной новостной информации	Logit	0.4632	0.6155	0.5394	-
	Random_forest	0.4696	0.5492	0.5094	-
	Bayes_Naive	0.4539	0.6028	0.5283	-
	Boosting	0.4468	0.5800	0.5134	-
	SVM-Linear	0.4374	0.5955	0.5164	-
	SVM-Poly	0.4350	0.6375	0.5362	-
	SVM-Radial	0.4292	0.6312	0.5302	-
Оба канала информации (текущая)	Logit	0.4652	0.6038	0.5345	-
	Random_forest	0.4232	0.6057	0.5145	-
	Bayes_Naive	0.4209	0.6323	0.5266	-
	Boosting	0.4446	0.5984	0.5215	-
	SVM-Linear	0.4359	0.6272	0.5315	-
	SVM-Poly	0.4415	0.6295	0.5355	-
	SVM-Radial	0.4607	0.5863	0.5235	-
Оба канала информации (новости лагированные)	Logit	0.4351	0.6238	0.5294	-
	Random_forest	0.4498	0.5929	0.5213	-
	Bayes_Naive	0.4817	0.5869	0.5343	-
	Boosting	0.4614	0.5933	0.5273	-
	SVM-Linear	0.4542	0.5964	0.5253	-
	SVM-Poly	0.4654	0.6012	0.5333	-
	SVM-Radial	0.4373	0.6529	0.5451	-

Таблица 11. Классификация магнитуды изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4522	0.5488	0.5005	-
	Random_forest	0.4752	0.5695	0.5223	-
	Bayes_Naive	0.4726	0.6336	0.5531	-
	Boosting	0.4899	0.6026	0.5462	-
	SVM-Linear	0.4756	0.5770	0.5263	-
	SVM-Poly	0.4795	0.5870	0.5333	-
	SVM-Radial	0.4929	0.6096	0.5513	-
Использование только новостной информации (текущая)	Logit	0.4395	0.6211	0.5303	-
	Random_forest	0.4276	0.6308	0.5292	-
	Bayes_Naive	0.4382	0.6106	0.5244	-
	Boosting	0.4624	0.6022	0.5323	-
	SVM-Linear	0.4210	0.6217	0.5213	-
	SVM-Poly	0.4444	0.6301	0.5372	-
	SVM-Radial	0.4569	0.6275	0.5422	-
Использование лагированной новостной информации	Logit	0.4686	0.6437	0.5561	-
	Random_forest	0.4202	0.6046	0.5124	-
	Bayes_Naive	0.4882	0.6302	0.5592	-
	Boosting	0.4611	0.6174	0.5392	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (текущая)	SVM-Linear	0.4852	0.6350	0.5601	-
	SVM-Poly	0.4756	0.6505	0.5631	-
	SVM-Radial	0.4730	0.6135	0.5432	-
	Logit	0.4379	0.6347	0.5363	-
	Random_forest	0.4642	0.6521	0.5581	-
	Bayes_Naive	0.4742	0.6420	0.5581	-
	Boosting	0.4922	0.6538	0.5730	-
	SVM-Linear	0.4252	0.6115	0.5184	-
	SVM-Poly	0.4629	0.6155	0.5392	-
	SVM-Radial	0.5147	0.6354	0.5750	+
Оба канала информации (новости лагированные)	Logit	0.4639	0.6582	0.5610	-
	Random_forest	0.4413	0.6294	0.5354	-
	Bayes_Naive	0.4987	0.6375	0.5681	-
	Boosting	0.5073	0.5713	0.5393	+
	SVM-Linear	0.4791	0.6371	0.5581	-
	SVM-Poly	0.4539	0.6642	0.5591	-
	SVM-Radial	0.4963	0.6021	0.5492	-

Таблица 12. Классификация направления доходности по важным для компании словам

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4740	0.6045	0.5393	-
	Random_forest	0.4542	0.5568	0.5055	-
	Bayes_Naive	0.4457	0.6090	0.5273	-
	Boosting	0.4277	0.6171	0.5224	-
	SVM-Linear	0.4897	0.5450	0.5174	-
	SVM-Poly	0.5183	0.5265	0.5224	+
	SVM-Radial	0.4236	0.5893	0.5065	-
Использование только новостной информации (текущая)	Logit	0.4902	0.5427	0.5164	-
	Random_forest	0.4836	0.5690	0.5263	-
	Bayes_Naive	0.4722	0.5487	0.5104	-
	Boosting	0.5191	0.5535	0.5363	+
	SVM-Linear	0.4887	0.5402	0.5144	-
	SVM-Poly	0.5179	0.5448	0.5313	+
	SVM-Radial	0.5095	0.5532	0.5313	+
Использование лагированной новостной информации	Logit	0.4874	0.5445	0.5160	-
	Random_forest	0.4800	0.5657	0.5229	-
	Bayes_Naive	0.4870	0.5467	0.5169	-
	Boosting	0.5035	0.5681	0.5358	+
	SVM-Linear	0.4871	0.5368	0.5119	-
	SVM-Poly	0.5048	0.5450	0.5249	+
	SVM-Radial	0.4866	0.5731	0.5298	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (текущая)	Logit	0.4678	0.5988	0.5333	-
	Random_forest	0.5136	0.5312	0.5224	+
	Bayes_Naive	0.4668	0.5522	0.5095	-
	Boosting	0.4494	0.6231	0.5363	-
	SVM-Linear	0.4903	0.5406	0.5154	-
	SVM-Poly	0.5179	0.5448	0.5313	+
	SVM-Radial	0.4639	0.5649	0.5144	-
Оба канала информации (новости лагированные)	Logit	0.4697	0.5900	0.5299	-
	Random_forest	0.5175	0.5303	0.5239	+
	Bayes_Naive	0.4870	0.5467	0.5169	-
	Boosting	0.4581	0.6018	0.5300	-
	SVM-Linear	0.4991	0.5288	0.5139	-
	SVM-Poly	0.5048	0.5450	0.5249	+
	SVM-Radial	0.4637	0.5562	0.5100	-

Таблица 13. Классификация масштаба доходности по важным для компании словам

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4522	0.5488	0.5005	-
	Random_forest	0.4752	0.5695	0.5223	-
	Bayes_Naive	0.4726	0.6336	0.5531	-
	Boosting	0.4899	0.6026	0.5462	-
	SVM-Linear	0.4756	0.5770	0.5263	-
	SVM-Poly	0.4795	0.5870	0.5333	-
	SVM-Radial	0.4929	0.6096	0.5513	-
Использование только новостной информации (текущая)	Logit	0.4640	0.5669	0.5154	-
	Random_forest	0.4751	0.5696	0.5223	-
	Bayes_Naive	0.4685	0.5484	0.5084	-
	Boosting	0.4842	0.5486	0.5164	-
	SVM-Linear	0.4662	0.5566	0.5114	-
	SVM-Poly	0.4866	0.5621	0.5243	-
	SVM-Radial	0.4717	0.5710	0.5214	-
Использование лагированной новостной информации	Logit	0.4866	0.5354	0.5110	-
	Random_forest	0.4575	0.5765	0.5170	-
	Bayes_Naive	0.4728	0.5352	0.5040	-
	Boosting	0.4812	0.5507	0.5160	-
	SVM-Linear	0.4723	0.5497	0.5110	-
	SVM-Poly	0.4842	0.5577	0.5210	-
	SVM-Radial	0.4649	0.5710	0.5180	-
Оба канала информации (текущая)	Logit	0.4646	0.6079	0.5363	-
	Random_forest	0.4843	0.5963	0.5403	-
	Bayes_Naive	0.4656	0.5592	0.5124	-
	Boosting	0.5181	0.6221	0.5701	+

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
	SVM-Linear	0.4679	0.5549	0.5114	-
	SVM-Poly	0.4866	0.5621	0.5243	-
	SVM-Radial	0.4993	0.6230	0.5611	-
Оба канала информации (новости ланированные)	Logit	0.4376	0.5806	0.5091	-
	Random_forest	0.4724	0.6013	0.5368	-
	Bayes_Naive	0.4570	0.5848	0.5209	-
	Boosting	0.4832	0.6403	0.5617	-
	SVM-Linear	0.4859	0.5421	0.5140	-
	SVM-Poly	0.4842	0.5577	0.5210	-
	SVM-Radial	0.4398	0.6301	0.5350	-



Таблица 1. ARIMA-модель для модели с применением сентимент-анализа

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ma1	0.2203	0.0309	7.1203	0.0000
intercept	− 0.0001	0.0008	− 0.0919	0.9268
sent_1	0.0001	0.0001	1.2091	0.2266
AIC	− 5 013			
BIC	− 4 993			
n.obs	1 016			

Таблица 2. Прогнозирование больших по модулю доходностей с использованием сентимент-анализа

Источник: Составлено автором

	int_a_lm	int_b_lm	cor_lm	pv_lm
volkw	− 0.1174	0.1504	0.0168	0.8373

Таблица 3. Прогнозирование доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	0.0709	0.2534	0.1636	0.0039
	Poly	0.0709	0.2534	0.1636	0.0039
	Radial	-0.0289	0.1578	0.0650	0.2545
Только сентимент в усеченной форме	Linear	-0.2254	-0.0412	-0.1344	0.0181
	Poly	-0.2379	-0.0545	-0.1475	0.0094
	Radial	-0.1210	0.0664	-0.0275	0.6300
Только сентимент в значимой форме	Linear	-0.0905	0.0970	0.0032	0.9547
	Poly	-0.0872	0.1003	0.0066	0.9086
	Radial	-0.1081	0.0793	-0.0145	0.7992
Финансы + усеченная форма сентимента	Linear	0.0310	0.2156	0.1244	0.0288
	Poly	0.0289	0.2136	0.1223	0.0316
	Radial	0.0091	0.1946	0.1028	0.0712
Финансы + значимая форма сентимента	Linear	0.0678	0.2505	0.1605	0.0047
	Poly	0.0682	0.2509	0.1609	0.0046
	Radial	0.0279	0.2127	0.1213	0.0330

Таблица 4. Прогнозирование больших доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	0.0291	0.2901	0.1625	0.0455
	Poly	0.0291	0.2901	0.1625	0.0455
	Radial	-0.0030	0.2603	0.1310	0.1078
Только сентимент в усеченной форме	Linear	-0.2075	0.0589	-0.0756	0.3545
	Poly	-0.1499	0.1180	-0.0162	0.8426
	Radial	-0.1417	0.1262	-0.0079	0.9232
Только сентимент в значимой форме	Linear	-0.1174	0.1504	0.0168	0.8373
	Poly	-0.1174	0.1504	0.0168	0.8373
	Radial	-0.2006	0.0661	-0.0685	0.4018
Финансы и усеченная форма сентимента	Linear	0.0066	0.2693	0.1404	0.0844
	Poly	0.0122	0.2744	0.1459	0.0730
	Radial	-0.0088	0.2549	0.1253	0.1241
Финансы и значимая форма сентимента	Linear	0.0222	0.2837	0.1556	0.0555
	Poly	0.0165	0.2785	0.1501	0.0649
	Radial	-0.0825	0.1846	0.0520	0.5248

Таблица 5. Классификация по направлению движения акций с использованием методов машинного обучения и сентимент-анализа

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5746	0.6463	0.6105	+
	Random_forest	0.4953	0.6250	0.5602	-
	Bayes_Naive	0.5153	0.6800	0.5977	+
	Boosting	0.5244	0.6373	0.5808	+
	SVM-Linear	0.5653	0.6398	0.6025	+
	SVM-Poly	0.5644	0.6664	0.6154	+
	SVM-Radial	0.5551	0.6460	0.6006	+
Использование только новостной информации (текущая и первый лаг)	Logit	0.5055	0.5557	0.5306	+
	Random_forest	0.4918	0.6027	0.5473	-
	Bayes_Naive	0.5047	0.5742	0.5394	+
	Boosting	0.5038	0.5967	0.5503	+
	SVM-Linear	0.5123	0.5251	0.5187	+
	SVM-Poly	0.5028	0.5426	0.5227	+
	SVM-Radial	0.5102	0.6158	0.5630	+
Использование лагированной новостной информации	Logit	0.4962	0.5453	0.5207	-
	Random_forest	0.3962	0.5682	0.4822	-
	Bayes_Naive	0.4736	0.5441	0.5089	-
	Boosting	0.4112	0.5533	0.4822	-
	SVM-Linear	0.5123	0.5251	0.5187	+

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
	SVM-Poly	0.4909	0.5664	0.5286	-
	SVM-Radial	0.4764	0.5512	0.5138	-
Оба канала информации (текущая и первый лаг)	Logit	0.5585	0.6564	0.6075	+
	Random_forest	0.5028	0.6488	0.5758	+
	Bayes_Naive	0.5195	0.6639	0.5917	+
	Boosting	0.5293	0.6522	0.5907	+
	SVM-Linear	0.5511	0.6521	0.6016	+
	SVM-Poly	0.5564	0.6684	0.6124	+
	SVM-Radial	0.5446	0.6624	0.6035	+
Оба канала информации (новости лагированные)	Logit	0.5585	0.6564	0.6075	+
	Random_forest	0.4935	0.6247	0.5591	-
	Bayes_Naive	0.4908	0.6176	0.5542	-
	Boosting	0.5432	0.6344	0.5888	+
	SVM-Linear	0.5506	0.6387	0.5947	+
	SVM-Poly	0.5372	0.6618	0.5995	+
	SVM-Radial	0.5120	0.6201	0.5661	+

Таблица 6. Классификация по размаху движения с использованием ML и сентиментов

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4253	0.6083	0.5168	-
	Random_forest	0.4263	0.5597	0.4930	-
	Bayes_Naive	0.4908	0.6255	0.5581	-
	Boosting	0.4759	0.6345	0.5552	-
	SVM-Linear	0.4701	0.6108	0.5404	-
	SVM-Poly	0.4725	0.6183	0.5454	-
	SVM-Radial	0.4919	0.6401	0.5660	-
Использование только новостной информации (текущая и первый лаг)	Logit	0.4573	0.5724	0.5149	-
	Random_forest	0.3930	0.6249	0.5090	-
	Bayes_Naive	0.4437	0.5541	0.4989	-
	Boosting	0.4318	0.5861	0.5089	-
	SVM-Linear	0.4536	0.5798	0.5167	-
	SVM-Poly	0.4575	0.5898	0.5236	-
	SVM-Radial	0.4364	0.5717	0.5040	-
Использование лагированной новостной информации	Logit	0.4452	0.5685	0.5068	-
	Random_forest	0.4292	0.5924	0.5108	-
	Bayes_Naive	0.4308	0.5417	0.4863	-
	Boosting	0.4178	0.5940	0.5059	-
	SVM-Linear	0.4430	0.5745	0.5088	-
	SVM-Poly	0.4669	0.5982	0.5325	-
	SVM-Radial	0.4711	0.6057	0.5384	-
Оба канала информации	Logit	0.4331	0.6044	0.5188	-
	Random_forest	0.4367	0.5611	0.4989	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
(текущая и первый лаг)	Bayes_Naive	0.4851	0.6272	0.5561	-
	Boosting	0.4729	0.6039	0.5384	-
	SVM-Linear	0.4305	0.6168	0.5236	-
	SVM-Poly	0.4919	0.6026	0.5473	-
	SVM-Radial	0.4852	0.6232	0.5542	-
Оба канала информации (новости ланированные)	Logit	0.4179	0.6116	0.5147	-
	Random_forest	0.4136	0.6512	0.5324	-
	Bayes_Naive	0.4727	0.6200	0.5463	-
	Boosting	0.4432	0.6376	0.5404	-
	SVM-Linear	0.4096	0.6278	0.5187	-
	SVM-Poly	0.4568	0.6162	0.5365	-
	SVM-Radial	0.4568	0.6478	0.5523	-

Таблица 7. ARIMA-модель для Фольксваген с выделением топигов по методу LDA

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ar1	0.2171	0.0306	7.0895	0.0000
intercept	0.0001	0.0011	0.0742	0.9408
topik_3_3	- 0.0134	0.0082	- 1.6310	0.1029
topik_4	0.0121	0.0066	1.8232	0.0683
AIC	- 5 014			
BIC	- 4 990			
n.obs	1 016			

Статистически значимыми являются словари, посвященные автомобильным авариям и судебным разбирательствам, а также информации о производстве автомобилей марки Фольксваген.

Таблица 8. Прогнозы доходностей с помощью модели с использованием LDA

Источник: составлено автором

	int_a_arima_dir	int_b_arima_dir	cor_arima_dir	pv_arima_dir
volkw	- 0.1159	0.1073	- 0.0044	0.9390

Таблица 9. Прогнозы дневных доходностей с помощью ML и LDA-словарей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Финансовая информация	Linear	0.0544	0.2716	0.1650	0.0036
	Poly	0.0553	0.2724	0.1658	0.0035
	Radial	0.0742	0.2899	0.1843	0.0011

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только текущие словари	Linear	-0.1790	0.0431	-0.0688	0.2281
	Poly	-0.1874	0.0345	-0.0774	0.1746
	Radial	-0.1804	0.0416	-0.0703	0.2180
Только лаги словарей	Linear	-0.1891	0.0326	-0.0792	0.1647
	Poly	-0.2032	0.0180	-0.0938	0.0998
	Radial	-0.0741	0.1487	0.0378	0.5085
2 источника – текущие словари	Linear	0.0444	0.2623	0.1552	0.0062
	Poly	0.0387	0.2569	0.1496	0.0084
	Radial	-0.0050	0.2156	0.1066	0.0612
2 источника – лаги словарей	Linear	0.0157	0.2353	0.1270	0.0256
	Poly	-0.0223	0.1991	0.0895	0.1163
	Radial	0.0066	0.2267	0.1181	0.0379

Таблица 10. Классификация направления изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5746	0.6463	0.6105	+
	Random_forest	0.4953	0.6250	0.5602	-
	Bayes_Naive	0.5153	0.6800	0.5977	+
	Boosting	0.5244	0.6373	0.5808	+
	SVM-Linear	0.5653	0.6398	0.6025	+
	SVM-Poly	0.5644	0.6664	0.6154	+
	SVM-Radial	0.5551	0.6460	0.6006	+
Использование только новостной информации (текущая)	Logit	0.4178	0.5764	0.4971	-
	Random_forest	0.4474	0.6236	0.5355	-
	Bayes_Naive	0.4084	0.5698	0.4891	-
	Boosting	0.4202	0.5522	0.4862	-
	SVM-Linear	0.4356	0.5822	0.5089	-
	SVM-Poly	0.5123	0.5251	0.5187	+
	SVM-Radial	0.4282	0.5898	0.5090	-
Использование лагированной новостной информации	Logit	0.4588	0.5392	0.4990	-
	Random_forest	0.4731	0.5861	0.5296	-
	Bayes_Naive	0.4103	0.5780	0.4941	-
	Boosting	0.4618	0.5521	0.5070	-
	SVM-Linear	0.4443	0.5399	0.4921	-
	SVM-Poly	0.5006	0.5959	0.5482	+
	SVM-Radial	0.4991	0.5640	0.5316	-
Оба канала информации (текущая)	Logit	0.5377	0.6534	0.5956	+
	Random_forest	0.5250	0.6309	0.5779	+
	Bayes_Naive	0.4597	0.6192	0.5394	-
	Boosting	0.5308	0.6448	0.5878	+
	SVM-Linear	0.5389	0.6169	0.5779	+
	SVM-Poly	0.4860	0.6620	0.5740	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (новости лагированные)	SVM-Radial	0.4951	0.6469	0.5710	-
	Logit	0.5286	0.6567	0.5926	+
	Random_forest	0.5247	0.6291	0.5769	+
	Bayes_Naive	0.4200	0.6253	0.5227	-
	Boosting	0.4845	0.6358	0.5601	-
	SVM-Linear	0.5122	0.6377	0.5750	+
	SVM-Poly	0.5048	0.6353	0.5700	+
	SVM-Radial	0.5069	0.6253	0.5661	+

Таблица 11. Классификация магнитуды изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4253	0.6083	0.5168	-
	Random_forest	0.4263	0.5597	0.4930	-
	Bayes_Naive	0.4908	0.6255	0.5581	-
	Boosting	0.4759	0.6345	0.5552	-
	SVM-Linear	0.4701	0.6108	0.5404	-
	SVM-Poly	0.4725	0.6183	0.5454	-
	SVM-Radial	0.4919	0.6401	0.5660	-
Использование только новостной информации (текущая)	Logit	0.4073	0.5927	0.5000	-
	Random_forest	0.3817	0.6262	0.5040	-
	Bayes_Naive	0.3925	0.5599	0.4762	-
	Boosting	0.3797	0.5471	0.4634	-
	SVM-Linear	0.3952	0.5634	0.4793	-
	SVM-Poly	0.4594	0.5879	0.5237	-
	SVM-Radial	0.4246	0.5853	0.5050	-
Использование лагированной новостной информации	Logit	0.4182	0.5797	0.4990	-
	Random_forest	0.4119	0.6336	0.5228	-
	Bayes_Naive	0.4379	0.5976	0.5178	-
	Boosting	0.4241	0.5620	0.4930	-
	SVM-Linear	0.4357	0.5624	0.4990	-
	SVM-Poly	0.4414	0.5645	0.5030	-
	SVM-Radial	0.4371	0.6004	0.5188	-
Оба канала информации (текущая)	Logit	0.4251	0.6047	0.5149	-
	Random_forest	0.4365	0.6148	0.5256	-
	Bayes_Naive	0.4731	0.6137	0.5434	-
	Boosting	0.4530	0.6101	0.5315	-
	SVM-Linear	0.4167	0.5835	0.5001	-
	SVM-Poly	0.4660	0.5951	0.5306	-
	SVM-Radial	0.4538	0.5756	0.5147	-
Оба канала информации (новости лагированные)	Logit	0.4368	0.5809	0.5089	-
	Random_forest	0.4385	0.6287	0.5336	-
	Bayes_Naive	0.4464	0.6443	0.5454	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
	Boosting	0.4355	0.6001	0.5178	-
	SVM-Linear	0.4086	0.6070	0.5078	-
	SVM-Poly	0.4362	0.5934	0.5148	-
	SVM-Radial	0.4423	0.6012	0.5217	-

Таблица 12. Классификация направления изменения доходности по важным для компании Фольксваген словам (TF-IDF)

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5746	0.6463	0.6105	+
	Random_forest	0.4953	0.6250	0.5602	-
	Bayes_Naive	0.5153	0.6800	0.5977	+
	Boosting	0.5244	0.6373	0.5808	+
	SVM-Linear	0.5653	0.6398	0.6025	+
	SVM-Poly	0.5644	0.6664	0.6154	+
	SVM-Radial	0.5551	0.6460	0.6006	+
Использование только новостной информации (текущая)	Logit	0.4883	0.5589	0.5236	-
	Random_forest	0.5123	0.5251	0.5187	+
	Bayes_Naive	0.5006	0.5369	0.5187	+
	Boosting	0.4865	0.5510	0.5187	-
	SVM-Linear	0.4866	0.5507	0.5187	-
	SVM-Poly	0.5107	0.5327	0.5217	+
	SVM-Radial	0.5082	0.5273	0.5177	+
Использование лагированной новостной информации	Logit	0.4763	0.5679	0.5221	-
	Random_forest	0.5133	0.5252	0.5192	+
	Bayes_Naive	0.5029	0.5337	0.5183	+
	Boosting	0.4766	0.5441	0.5104	-
	SVM-Linear	0.4668	0.5696	0.5182	-
	SVM-Poly	0.5119	0.5325	0.5222	+
	SVM-Radial	0.5133	0.5252	0.5192	+
Оба канала информации (текущая)	Logit	0.5531	0.6579	0.6055	+
	Random_forest	0.5287	0.5995	0.5641	+
	Bayes_Naive	0.5006	0.5369	0.5187	+
	Boosting	0.5458	0.6495	0.5976	+
	SVM-Linear	0.4864	0.5667	0.5265	-
	SVM-Poly	0.5107	0.5327	0.5217	+
	SVM-Radial	0.5285	0.6373	0.5829	+
Оба канала информации (новости лагированные)	Logit	0.5371	0.6692	0.6031	+
	Random_forest	0.4877	0.6476	0.5677	-
	Bayes_Naive	0.5029	0.5337	0.5183	+
	Boosting	0.5125	0.6602	0.5864	+
	SVM-Linear	0.4726	0.5776	0.5251	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
	SVM-Poly	0.5119	0.5325	0.5222	+
	SVM-Radial	0.4997	0.6475	0.5736	-

Таблица 13. Классификация масштаба доходности по важным для компании словам

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4253	0.6083	0.5168	-
	Random_forest	0.4263	0.5597	0.4930	-
	Bayes_Naive	0.4908	0.6255	0.5581	-
	Boosting	0.4759	0.6345	0.5552	-
	SVM-Linear	0.4701	0.6108	0.5404	-
	SVM-Poly	0.4725	0.6183	0.5454	-
	SVM-Radial	0.4919	0.6401	0.5660	-
Использование только новостной информации (текущая)	Logit	0.4740	0.5596	0.5168	-
	Random_forest	0.4657	0.5581	0.5119	-
	Bayes_Naive	0.4947	0.5289	0.5118	-
	Boosting	0.4572	0.5605	0.5089	-
	SVM-Linear	0.4596	0.5502	0.5049	-
	SVM-Poly	0.4825	0.5174	0.5000	-
	SVM-Radial	0.4669	0.5450	0.5060	-
Использование лагированной новостной информации	Logit	0.4963	0.5363	0.5163	-
	Random_forest	0.4666	0.5541	0.5103	-
	Bayes_Naive	0.4751	0.5712	0.5232	-
	Boosting	0.4680	0.5549	0.5114	-
	SVM-Linear	0.4868	0.5458	0.5163	-
	SVM-Poly	0.4713	0.5336	0.5025	-
	SVM-Radial	0.4700	0.5429	0.5065	-
Оба канала информации (текущая)	Logit	0.4678	0.6111	0.5395	-
	Random_forest	0.4760	0.5830	0.5295	-
	Bayes_Naive	0.4952	0.5323	0.5138	-
	Boosting	0.4856	0.6268	0.5562	-
	SVM-Linear	0.4558	0.5520	0.5039	-
	SVM-Poly	0.4114	0.5962	0.5038	-
	SVM-Radial	0.4454	0.6611	0.5532	-
Оба канала информации (новости ланированные)	Logit	0.5152	0.5924	0.5538	+
	Random_forest	0.4801	0.5821	0.5311	-
	Bayes_Naive	0.4639	0.5923	0.5281	-
	Boosting	0.4628	0.6585	0.5607	-
	SVM-Linear	0.4873	0.5434	0.5153	-
	SVM-Poly	0.4630	0.5399	0.5014	-
	SVM-Radial	0.4679	0.6691	0.5685	-

Результаты проведенного анализа по компании BMW



Таблица 1. ARIMA-модель для модели с применением сентимент-анализа

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ar1	0.2119	0.0306	6.9141	0.0000
intercept	0.0004	0.0006	0.5995	0.5489
sent_2	− 0.0002	0.0001	− 1.4572	0.1451
AIC	− 5 464			
BIC	− 5 445			
n.obs	1 016			

Таблица 2. Прогнозирование больших по модулю доходностей с использованием сентимент-анализа

Источник: Составлено автором

	int_a_lm	int_b_lm	cor_lm	pv_lm
bmw	− 0.1072	0.1605	0.0272	0.7397

Таблица 3. Прогнозирование доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	0.0669	0.2497	0.1596	0.0049
	Poly	0.0626	0.2456	0.1554	0.0062
	Radial	0.0672	0.2499	0.1599	0.0048
Только сентимент в усеченной форме	Linear	-0.0885	0.0990	0.0053	0.9255
	Poly	-0.0848	0.1027	0.0091	0.8740
	Radial	-0.0632	0.1241	0.0307	0.5906
Только сентимент в значимой форме	Linear	-0.0988	0.0887	-0.0051	0.9284
	Poly	-0.0601	0.1272	0.0339	0.5533
	Radial	-0.0554	0.1319	0.0386	0.4993
Финансы + усеченная форма сентимента	Linear	0.0625	0.2455	0.1553	0.0062
	Poly	0.0650	0.2479	0.1578	0.0054
	Radial	0.0498	0.2335	0.1429	0.0119
Финансы + значимая форма сентимента	Linear	0.0416	0.2257	0.1348	0.0177
	Poly	0.0200	0.2051	0.1135	0.0461
	Radial	0.0236	0.2085	0.1171	0.0397

Таблица 4. Прогнозирование больших доходностей с использованием ML-моделей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только финансы	Linear	0.0197	0.2814	0.1533	0.0594
	Poly	0.0128	0.2750	0.1464	0.0718
	Radial	0.0153	0.2774	0.1490	0.0670
Только сентимент в усеченной форме	Linear	-0.1121	0.1557	0.0222	0.7860
	Poly	-0.1088	0.1589	0.0255	0.7554
	Radial	-0.1182	0.1496	0.0160	0.8450
Только сентимент в значимой форме	Linear	-0.0456	0.2202	0.0889	0.2761
	Poly	-0.0410	0.2246	0.0935	0.2521
	Radial	-0.2365	0.0284	-0.1059	0.1941
Финансы и усеченная форма сентимента	Linear	0.0468	0.3061	0.1796	0.0269
	Poly	0.0320	0.2927	0.1652	0.0420
	Radial	0.0166	0.2785	0.1502	0.0648
Финансы и значимая форма сентимента	Linear	0.0411	0.3010	0.1741	0.0320
	Poly	0.0205	0.2821	0.1540	0.0582
	Radial	0.0124	0.2746	0.1461	0.0725

Таблица 5. Классификация по направлению движения акций с использованием методов машинного обучения и сентимент-анализа

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5461	0.6512	0.5987	+
	Random_forest	0.4457	0.5995	0.5226	-
	Bayes_Naive	0.5383	0.6236	0.5809	+
	Boosting	0.5537	0.6199	0.5868	+
	SVM-Linear	0.5147	0.6531	0.5839	+
	SVM-Poly	0.5381	0.6473	0.5927	+
	SVM-Radial	0.5337	0.6260	0.5798	+
Использование только новостной информации (текущая и первый лаг)	Logit	0.4893	0.5698	0.5295	-
	Random_forest	0.4307	0.6047	0.5177	-
	Bayes_Naive	0.3933	0.5930	0.4931	-
	Boosting	0.4495	0.5956	0.5226	-
	SVM-Linear	0.5229	0.5362	0.5296	+
	SVM-Poly	0.5229	0.5362	0.5296	+
	SVM-Radial	0.4621	0.6010	0.5315	-
Использование лагированной новостной информации	Logit	0.4849	0.5586	0.5217	-
	Random_forest	0.4302	0.5932	0.5117	-
	Bayes_Naive	0.4380	0.6013	0.5197	-
	Boosting	0.4416	0.5819	0.5117	-
	SVM-Linear	0.5229	0.5362	0.5296	+
	SVM-Poly	0.5229	0.5362	0.5296	+

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (текущая и первый лаг)	SVM-Radial	0.4563	0.6400	0.5482	-
	Logit	0.5352	0.6504	0.5928	+
	Random_forest	0.4954	0.5952	0.5453	-
	Bayes_Naive	0.5220	0.6004	0.5612	+
	Boosting	0.5507	0.6327	0.5917	+
	SVM-Linear	0.5205	0.6453	0.5829	+
	SVM-Poly	0.5343	0.6453	0.5898	+
	SVM-Radial	0.5407	0.6270	0.5838	+
Оба канала информации (новости лагированные)	Logit	0.5371	0.6622	0.5997	+
	Random_forest	0.4881	0.6577	0.5729	-
	Bayes_Naive	0.4955	0.6286	0.5621	-
	Boosting	0.5543	0.6430	0.5987	+
	SVM-Linear	0.5109	0.6589	0.5849	+
	SVM-Poly	0.5271	0.6663	0.5967	+
	SVM-Radial	0.5075	0.6583	0.5829	+

Таблица 6. Классификация по размаху движения с использованием ML и сентиментов

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4219	0.6255	0.5237	-
	Random_forest	0.4726	0.6301	0.5514	-
	Bayes_Naive	0.5288	0.6330	0.5809	+
	Boosting	0.5256	0.6500	0.5878	+
	SVM-Linear	0.4724	0.5789	0.5257	-
	SVM-Poly	0.5102	0.5983	0.5543	+
	SVM-Radial	0.5075	0.6564	0.5819	+
Использование только новостной информации (текущая и первый лаг)	Logit	0.4496	0.5585	0.5040	-
	Random_forest	0.4490	0.6003	0.5247	-
	Bayes_Naive	0.4446	0.5553	0.5000	-
	Boosting	0.4413	0.5352	0.4882	-
	SVM-Linear	0.4407	0.5536	0.4971	-
	SVM-Poly	0.4658	0.5383	0.5021	-
	SVM-Radial	0.4560	0.5834	0.5197	-
Использование лагированной новостной информации	Logit	0.4038	0.5529	0.4784	-
	Random_forest	0.4686	0.5906	0.5296	-
	Bayes_Naive	0.4589	0.5787	0.5188	-
	Boosting	0.4598	0.5482	0.5040	-
	SVM-Linear	0.4045	0.5663	0.4854	-
	SVM-Poly	0.4513	0.5487	0.5000	-
	SVM-Radial	0.4894	0.5521	0.5207	-
Оба канала информации	Logit	0.4278	0.5862	0.5070	-
	Random_forest	0.4901	0.6204	0.5553	-
	Bayes_Naive	0.4883	0.6418	0.5651	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
(текущая и первый лаг)	Boosting	0.5233	0.6542	0.5887	+
	SVM-Linear	0.4593	0.5566	0.5079	-
	SVM-Poly	0.4865	0.6300	0.5583	-
	SVM-Radial	0.4801	0.6661	0.5731	-
Оба канала информации (новости ланированные)	Logit	0.3955	0.6104	0.5030	-
	Random_forest	0.5022	0.6479	0.5751	+
	Bayes_Naive	0.4945	0.6181	0.5563	-
	Boosting	0.5089	0.6273	0.5681	+
	SVM-Linear	0.4655	0.5781	0.5218	-
	SVM-Poly	0.4815	0.6489	0.5652	-
	SVM-Radial	0.5086	0.6552	0.5819	+

Таблица 7. ARIMA-модель для БМВ с выделением топиков по методу LDA

Источник: составлено автором

	estimate	sd	t_stat	pi_val
ar1	0.2067	0.0307	6.7257	0.0000
intercept	- 0.0001	0.0008	- 0.1387	0.8897
topik_5_2	- 0.0119	0.0059	- 2.0256	0.0428
topik_7_2	0.0102	0.0058	1.7512	0.0799
topik_7_3	0.0114	0.0058	1.9597	0.0500
AIC	- 5 467			
BIC	- 5 437			
n.obs	1 016			

Статистически значимыми являются тематические группы, затрагивающие судебные разбирательства и вопросы, связанные с производством автомобилей, функционированием заводов компании.

Таблица 8. Прогнозы доходностей с помощью модели с использованием LDA

Источник: составлено автором

	int_a_arima_dir	int_b_arima_dir	cor_arima_dir	pv_arima_dir
bmw	- 0.1228	0.1004	-0.0114	0.8425

Таблица 9. Прогнозы дневных доходностей с помощью ML и LDA-словарей

Источник: составлено автором

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Финансовая информация	Linear	0.0554	0.2725	0.1660	0.0034
	Poly	0.0305	0.2493	0.1416	0.0127
	Radial	0.0214	0.2406	0.1326	0.0197

	kernel_svm	int_a_svm	int_b_svm	cor_svm	pv_svm
Только текущие словари	Linear	-0.1837	0.0382	-0.0736	0.1967
	Poly	-0.1825	0.0395	-0.0724	0.2043
	Radial	-0.1087	0.1145	0.0029	0.9593
Только лаги словарей	Linear	-0.2197	0.0008	-0.1108	0.0516
	Poly	-0.1930	0.0286	-0.0832	0.1445
	Radial	-0.0895	0.1336	0.0223	0.6957
2 источника – текущие словари	Linear	0.0138	0.2335	0.1252	0.0278
	Poly	0.0072	0.2272	0.1186	0.0371
	Radial	0.0056	0.2258	0.1172	0.0396
2 источника – лаги словарей	Linear	-0.0146	0.2065	0.0971	0.0883
	Poly	-0.0363	0.1856	0.0756	0.1852
	Radial	-0.0569	0.1656	0.0550	0.3353

Таблица 10. Классификация направления изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5461	0.6512	0.5987	+
	Random_forest	0.4457	0.5995	0.5226	-
	Bayes_Naive	0.5383	0.6236	0.5809	+
	Boosting	0.5537	0.6199	0.5868	+
	SVM-Linear	0.5147	0.6531	0.5839	+
	SVM-Poly	0.5381	0.6473	0.5927	+
	SVM-Radial	0.5337	0.6260	0.5798	+
Использование только новостной информации (текущая)	Logit	0.4477	0.5681	0.5079	-
	Random_forest	0.4409	0.5728	0.5069	-
	Bayes_Naive	0.4777	0.5973	0.5375	-
	Boosting	0.4242	0.6151	0.5197	-
	SVM-Linear	0.4603	0.5773	0.5188	-
	SVM-Poly	0.5241	0.5371	0.5306	+
	SVM-Radial	0.4654	0.5603	0.5128	-
Использование лагированной новостной информации	Logit	0.4288	0.5853	0.5071	-
	Random_forest	0.3932	0.6090	0.5011	-
	Bayes_Naive	0.4372	0.5826	0.5099	-
	Boosting	0.4529	0.5867	0.5198	-
	SVM-Linear	0.4477	0.5464	0.4971	-
	SVM-Poly	0.5229	0.5362	0.5296	+
	SVM-Radial	0.5202	0.5371	0.5286	+
Оба канала информации (текущая)	Logit	0.5240	0.6694	0.5967	+
	Random_forest	0.4959	0.6440	0.5699	-
	Bayes_Naive	0.4757	0.6052	0.5404	-
	Boosting	0.5017	0.6304	0.5661	+
	SVM-Linear	0.5027	0.6217	0.5622	+
	SVM-Poly	0.5339	0.6378	0.5858	+

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Оба канала информации (новости лагированные)	SVM-Radial	0.5100	0.6321	0.5710	+
	Logit	0.5534	0.6597	0.6065	+
	Random_forest	0.5041	0.6559	0.5800	+
	Bayes_Naive	0.4492	0.5784	0.5138	-
	Boosting	0.5101	0.6712	0.5907	+
	SVM-Linear	0.5142	0.6793	0.5968	+
	SVM-Poly	0.5224	0.6612	0.5918	+
	SVM-Radial	0.5049	0.6294	0.5671	+

Таблица 11. Классификация магнитуды изменения доходностей на LDA-словарях

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4219	0.6255	0.5237	-
	Random_forest	0.4726	0.6301	0.5514	-
	Bayes_Naive	0.5288	0.6330	0.5809	+
	Boosting	0.5256	0.6500	0.5878	+
	SVM-Linear	0.4724	0.5789	0.5257	-
	SVM-Poly	0.5102	0.5983	0.5543	+
	SVM-Radial	0.5075	0.6564	0.5819	+
Использование только новостной информации (текущая)	Logit	0.4662	0.5772	0.5217	-
	Random_forest	0.4651	0.5623	0.5137	-
	Bayes_Naive	0.4742	0.5712	0.5227	-
	Boosting	0.4949	0.5859	0.5404	-
	SVM-Linear	0.4519	0.5936	0.5227	-
	SVM-Poly	0.4679	0.5934	0.5307	-
	SVM-Radial	0.4591	0.6022	0.5306	-
Использование лагированной новостной информации	Logit	0.4165	0.6012	0.5088	-
	Random_forest	0.4322	0.6096	0.5209	-
	Bayes_Naive	0.4359	0.6035	0.5197	-
	Boosting	0.4256	0.6475	0.5365	-
	SVM-Linear	0.4242	0.6074	0.5158	-
	SVM-Poly	0.4318	0.6237	0.5277	-
	SVM-Radial	0.4351	0.6182	0.5266	-
Оба канала информации (текущая)	Logit	0.4542	0.6013	0.5277	-
	Random_forest	0.5035	0.6445	0.5740	+
	Bayes_Naive	0.5324	0.6254	0.5789	+
	Boosting	0.4868	0.6631	0.5749	-
	SVM-Linear	0.4489	0.5967	0.5228	-
	SVM-Poly	0.4709	0.6339	0.5524	-
	SVM-Radial	0.5207	0.6431	0.5819	+
Оба канала информации (новости лагированные)	Logit	0.4099	0.6038	0.5069	-
	Random_forest	0.5123	0.6320	0.5721	+
	Bayes_Naive	0.4299	0.6448	0.5373	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
	Boosting	0.5110	0.6548	0.5829	+
	SVM-Linear	0.4354	0.6179	0.5266	-
	SVM-Poly	0.4382	0.6172	0.5277	-
	SVM-Radial	0.4680	0.6210	0.5445	-

Таблица 12. Классификация направления изменения доходности по важным для компании БМВ словам (TF-IDF)

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.5461	0.6512	0.5987	+
	Random_forest	0.4457	0.5995	0.5226	-
	Bayes_Naive	0.5383	0.6236	0.5809	+
	Boosting	0.5537	0.6199	0.5868	+
	SVM-Linear	0.5147	0.6531	0.5839	+
	SVM-Poly	0.5381	0.6473	0.5927	+
	SVM-Radial	0.5337	0.6260	0.5798	+
Использование только новостной информации (текущая)	Logit	0.5040	0.5789	0.5414	+
	Random_forest	0.4986	0.5685	0.5335	-
	Bayes_Naive	0.4579	0.5795	0.5187	-
	Boosting	0.4953	0.5600	0.5276	-
	SVM-Linear	0.4982	0.5729	0.5355	-
	SVM-Poly	0.5197	0.5454	0.5325	+
	SVM-Radial	0.4996	0.5675	0.5335	-
Использование лагированной новостной информации	Logit	0.4950	0.5929	0.5439	-
	Random_forest	0.4790	0.6009	0.5399	-
	Bayes_Naive	0.4404	0.5705	0.5054	-
	Boosting	0.4939	0.5663	0.5301	-
	SVM-Linear	0.4898	0.5763	0.5331	-
	SVM-Poly	0.5187	0.5455	0.5321	+
	SVM-Radial	0.4912	0.5867	0.5390	-
Оба канала информации (текущая)	Logit	0.5303	0.6612	0.5958	+
	Random_forest	0.4919	0.5890	0.5404	-
	Bayes_Naive	0.4553	0.5919	0.5236	-
	Boosting	0.5469	0.6445	0.5957	+
	SVM-Linear	0.4973	0.5757	0.5365	-
	SVM-Poly	0.5197	0.5454	0.5325	+
	SVM-Radial	0.5052	0.6054	0.5553	+
Оба канала информации (новости лагированные)	Logit	0.5384	0.6582	0.5983	+
	Random_forest	0.4598	0.5985	0.5291	-
	Bayes_Naive	0.4335	0.6109	0.5222	-
	Boosting	0.5276	0.6393	0.5834	+
	SVM-Linear	0.4918	0.5842	0.5380	-

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
	SVM-Poly	0.5187	0.5455	0.5321	+
	SVM-Radial	0.4813	0.6245	0.5529	-

Таблица 13. Классификация масштаба доходности по важным для компании словам

Источник: составлено автором

	ml_type	ml_sd_a	ml_sd_b	ml_accuracy	ml_better
Использование только финансовой информации	Logit	0.4219	0.6255	0.5237	-
	Random_forest	0.4726	0.6301	0.5514	-
	Bayes_Naive	0.5288	0.6330	0.5809	+
	Boosting	0.5256	0.6500	0.5878	+
	SVM-Linear	0.4724	0.5789	0.5257	-
	SVM-Poly	0.5102	0.5983	0.5543	+
	SVM-Radial	0.5075	0.6564	0.5819	+
Использование только новостной информации (текущая)	Logit	0.4433	0.5253	0.4843	-
	Random_forest	0.4530	0.5393	0.4961	-
	Bayes_Naive	0.4759	0.5260	0.5010	-
	Boosting	0.4541	0.5421	0.4981	-
	SVM-Linear	0.4381	0.5246	0.4814	-
	SVM-Poly	0.4712	0.5327	0.5020	-
	SVM-Radial	0.4526	0.5436	0.4981	-
Использование лагированной новостной информации	Logit	0.4385	0.5190	0.4788	-
	Random_forest	0.4787	0.5124	0.4956	-
	Bayes_Naive	0.4576	0.5374	0.4975	-
	Boosting	0.4501	0.5431	0.4966	-
	SVM-Linear	0.4408	0.5187	0.4798	-
	SVM-Poly	0.4845	0.5146	0.4995	-
	SVM-Radial	0.4714	0.5217	0.4966	-
Оба канала информации (текущая)	Logit	0.4056	0.6024	0.5040	-
	Random_forest	0.5033	0.6172	0.5603	+
	Bayes_Naive	0.4558	0.5697	0.5127	-
	Boosting	0.5205	0.6551	0.5878	+
	SVM-Linear	0.4392	0.5216	0.4804	-
	SVM-Poly	0.4712	0.5327	0.5020	-
	SVM-Radial	0.4796	0.6626	0.5711	-
Оба канала информации (новости ланированные)	Logit	0.4015	0.6195	0.5105	-
	Random_forest	0.4780	0.6355	0.5568	-
	Bayes_Naive	0.4657	0.5471	0.5064	-
	Boosting	0.4910	0.6879	0.5895	-
	SVM-Linear	0.4458	0.5038	0.4748	-
	SVM-Poly	0.4845	0.5146	0.4995	-
	SVM-Radial	0.4872	0.6522	0.5697	-