

Тема: Моделі для прогнозування трафіку у сфері інтернет-реклами

Виконав: студент групи КА-44

Артьомов Артем

Науковий керівник: д.т.н., професор Бідюк П.І.

Об'єкт дослідження. Предмет дослідження

- Об'єкт дослідження: трафік як поняття, належне до сфери інтернет-маркетингу, що має під собою на увазі потік відвідувачів на певні елементи інтернет-мережі (сайти, портали, сервіси) або, інакше кажучи, їх відвідуваність.
- Предметом дослідження даної роботи є методи моделювання і прогнозування певних процесів сфери інтернет-маркетингу.

- Основною проблемою усіх програмних продуктів, якими у сьогоденні користуються фахівці з медіапланування, є те, що в роботі можуть бути застосовані лише фактичні дані і лише з запізненням в два-чотири тижні.
- Результатами непередбачуваних вагомих змін у динаміці трафіку на певних ресурсах можуть бути:
 1. Недосягнення поставлених клієнтом KPI, невиконання цілей рекламних кампаній;
 2. Неповний продаж наявного інструментарію паблішером, втрата можливого прибутку;
 3. Перепродаж наявного інструментарію, неможливість задовольнити потреби усіх рекламодавців.

Постановка задачі

- Метою даної роботи є побудова моделі або групи моделей для подальшого отримання якісного прогнозу трафіку у сфері Інтернет-реклами.
- Необхідно застосувати і порівняти різні методи вирішення даної задачі.
- Необхідно перевірити наявність і за необхідності врахувати вплив таких важливих факторів, як погодні умови, державні свята та головні події року.

Модель є особливо корисною для фахівців з планування рекламних кампанії, для фахівців з побудови річних стратегій. Також вона може бути використана при створенні і просуванні власного сайту.

Модель може бути застосована для вдосконалення таких програмних забезпечень з моніторингу та аналізу інтернет-трафіку, як InMind, gemiusAudience.

Існуючі програми, що можуть вирішувати подібні задачі



GMDH Shell

Вихідні дані

	A	B	C	D	E	F	G	H	I	J	K
1	Ресурс	Недели	Охват %	Доля ЦА в сайте %	Аффинити индекс посещения	Среднее количество просмотров на посетителя	Охват	Доля аудитории %	Просмотры страниц %	Доля просмотров страниц %	Доля просмотра страниц ЦА в сайте %
12	isport.ua	1-2014	0,0128	0,2598	190,21		9 30423	0,0153	0,183	0,0001	0,183
13	isport.ua	2-2014	0,0162	0,3052	226,61		15 38584	0,0198	0,3352	0,0002	0,3352
14	isport.ua	3-2014	0,0183	0,4889	362,43		7 43532	0,0233	0,2399	0,0001	0,2399
15	isport.ua	4-2014	0,0123	0,3437	253,59		18 29954	0,016	0,382	0,0002	0,382
16	isport.ua	5-2014	0,0216	0,416	290		13 56240	0,0266	0,4261	0,0003	0,4261
59	isport.ua	6-2014	0,0139	0,2726	192,32		24 36104	0,018	0,3271	0,0004	0,3271
70	isport.ua	7-2014	0,0081	0,1224	87,54		19 21118	0,0112	0,3472	0,0002	0,3472
91	isport.ua	9-2014	0,0062	0,152	106,05		10 15758	0,0075	0,1539	0,0001	0,1539
92	isport.ua	10-2014	0,0081	0,3426	246,07		11 20482	0,0107	0,304	0,0001	0,304
93	isport.ua	11-2014	0,0077	0,1577	114,36		9 19470	0,0106	0,1792	0,0001	0,1792

- Умовно вихідні дані можна розділити на дві частини: основні змінні і допоміжні індикатори.
- На малюнку вище наведені основні змінні.

Деякі пояснення

Охват

Охват ресурса в абсолютных значениях. Количество человек из целевой аудитории, посетивших ресурс хотя бы один раз.

$$\text{Охват} = (\text{Охват\%} \times \text{Общее количество человек в целевой аудитории}) \div 100$$

Аффинити индекс посещения

Индекс сходства посещения. Отношение доли целевой аудитории в сайте к доле представителей целевой аудитории, ПОСЕТИВШИХ НЕ МЕНЕЕ ОДНОГО РЕСУРСА в анализируемом периоде.

$$\text{Аффинити индекс посещения} = \text{Доля ЦА в сайте\%} \div \text{Доля целевой аудитории среди посетивших в анализируемом периоде хотя бы один раз любой из ресурсов\%}$$

Доля аудитории %

Доля ресурса в общем пользовании ресурсов. Доля людей в целевой аудитории, которые посещали данный ресурс, по крайней мере, один раз за анализируемый период, среди ПОСЕТИВШИХ НЕ МЕНЕЕ ОДНОГО РЕСУРСА в анализируемом периоде.

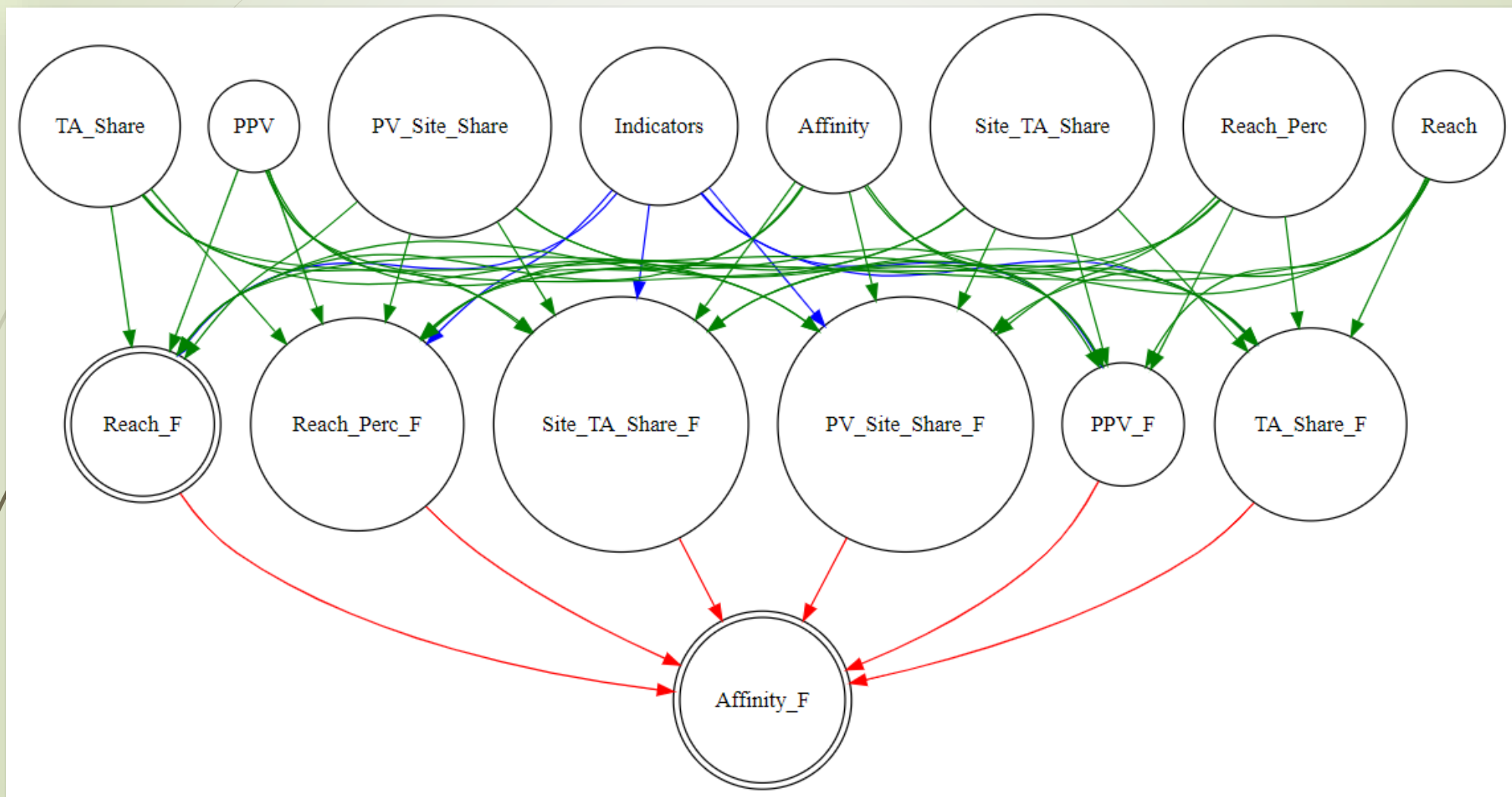
$$\text{Доля аудитории \%} = (\text{Количество человек в целевой аудитории, которые в анализируемом периоде посещали ресурс один раз и более} \div \text{Количество человек в целевой аудитории, которые в анализируемом периоде посещали хотя бы один раз любой из ресурсов}) \times 100$$

Вихідні дані

L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
Новый год	Рождество	8 марта	Пасха	День труда	День победы	Троица	День Конституции	День Независимости	День Защитника	Евровидение	ЧМ	ЧЕ	ЛЧ	Олимпиада	Погода
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,15
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4,6375
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,475
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-10,4625
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-14,75
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-10,225
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8875
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2,7625
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,9875

- Усі офіційні свята та найважливіші події року були розділені на окремі ряди, що містять лише «0» (подія відсутня) та «1» (подія відбулася). Це було зроблено для того, щоб за потреби їх можна було легко включати або виключати з моделей.

Структура моделі



Результати роботи у системі EViews. Числовий опис моделей

Dependent Variable: REACH

Method: Least Squares

Date: 05/22/18 Time: 19:45

Sample (adjusted): 11/03/2014 2/19/2018

Included observations: 172 after adjustments

REACH=C(1)+C(2)*REACH(-1)+C(3)*REACH(-21)+C(4)*REACH(-44)+C(5)
*AFFINITY+C(6)*TA_SHARE+C(7)*REACH_PERC+C(8)*CL+C(9)*EUC
+C(10)*MARCH+C(11)*EASTER+C(12)*CONSTITUTION_DAY+C(13)
*DEFEND_DAY+C(14)*NEW_YEAR+C(15)*CHRISTMAS+C(16)
*WEATHER

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-585.1702	551.5903	-1.060878	0.2904
C(2)	0.264317	0.029602	8.928893	0.0000
C(3)	0.024032	0.010955	2.193653	0.0297
C(4)	-0.002964	0.010971	-0.270153	0.7874
C(5)	-8.690092	2.264248	-3.837959	0.0002
C(6)	-565368.0	69377.22	-8.149188	0.0000
C(7)	3295600.	47120.55	69.93976	0.0000
C(8)	-359.7035	666.6799	-0.539545	0.5903
C(9)	-486.7065	899.7183	-0.540954	0.5893
C(10)	-864.7367	1120.885	-0.771477	0.4416
C(11)	-684.9346	798.8937	-0.857354	0.3926
C(12)	292.8916	1145.187	0.255759	0.7985
C(13)	828.3280	1088.580	0.760925	0.4479
C(14)	515.6324	868.4840	0.593716	0.5536
C(15)	823.8103	992.8370	0.829754	0.4079
C(16)	16.04238	15.00500	1.069136	0.2867

R-squared	0.984012	Mean dependent var	24437.44
Adjusted R-squared	0.982475	S.D. dependent var	13933.30
S.E. of regression	1844.523	Akaike info criterion	17.96624
Sum squared resid	5.31E+08	Schwarz criterion	18.25903
Log likelihood	-1529.096	Hannan-Quinn criter.	18.08503
F-statistic	640.0963	Durbin-Watson stat	1.419846
Prob(F-statistic)	0.000000		

Dependent Variable: AFFINITY

Method: Least Squares

Date: 05/22/18 Time: 22:32

Sample (adjusted): 5/05/2014 2/19/2018

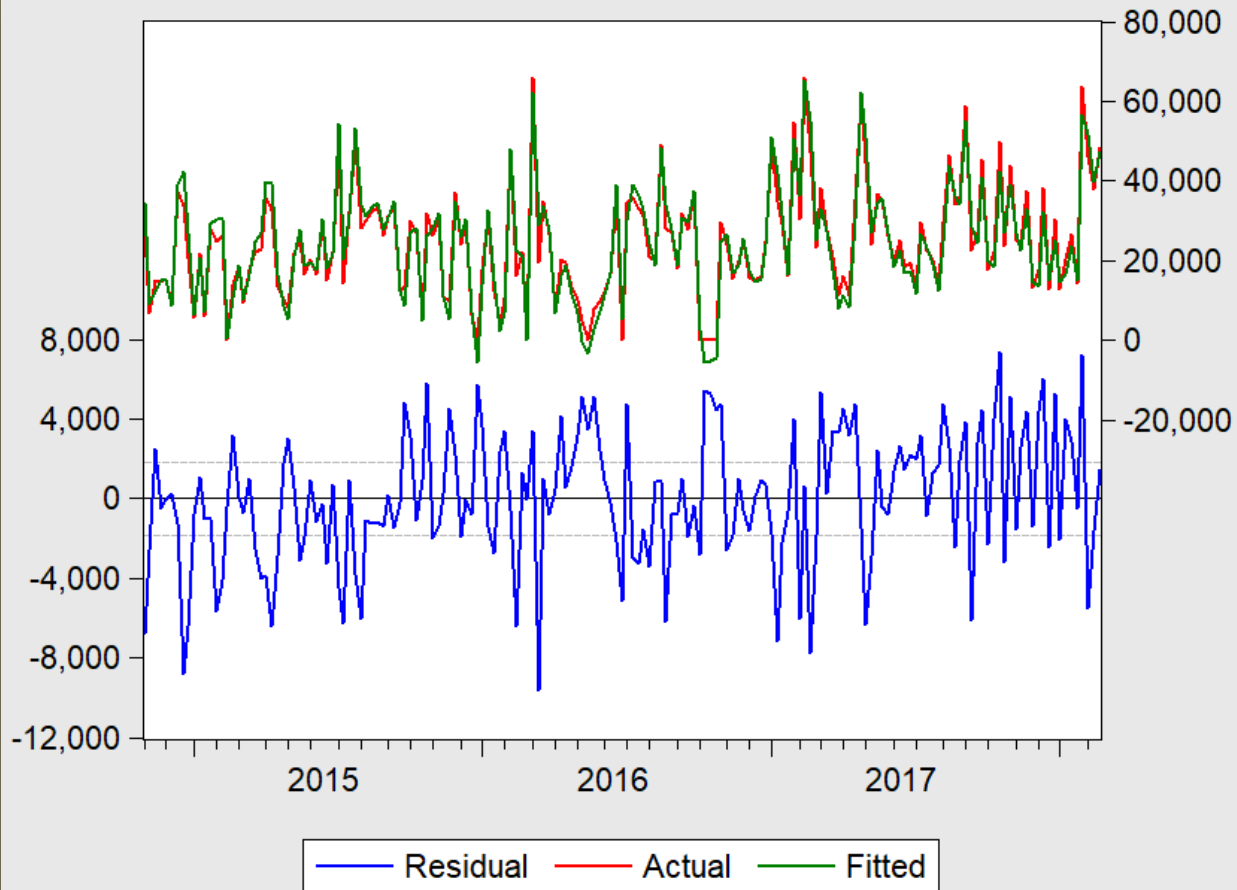
Included observations: 199 after adjustments

AFFINITY=C(1)+C(2)*AFFINITY(-1)+C(3)*AFFINITY(-5)+C(4)*AFFINITY(-6)
+C(5)*AFFINITY(-9)+C(6)*AFFINITY(-18)+C(7)*REACH_FORECAST
+C(8)*REACH_PERC_FORECAST+C(9)*PPV_FORECAST+C(10)
*PV_SITE_SHARE_FORECAST+C(11)*TA_SHARE_FORECAST+C(12)
*SITE_TA_SHARE_FORECAST

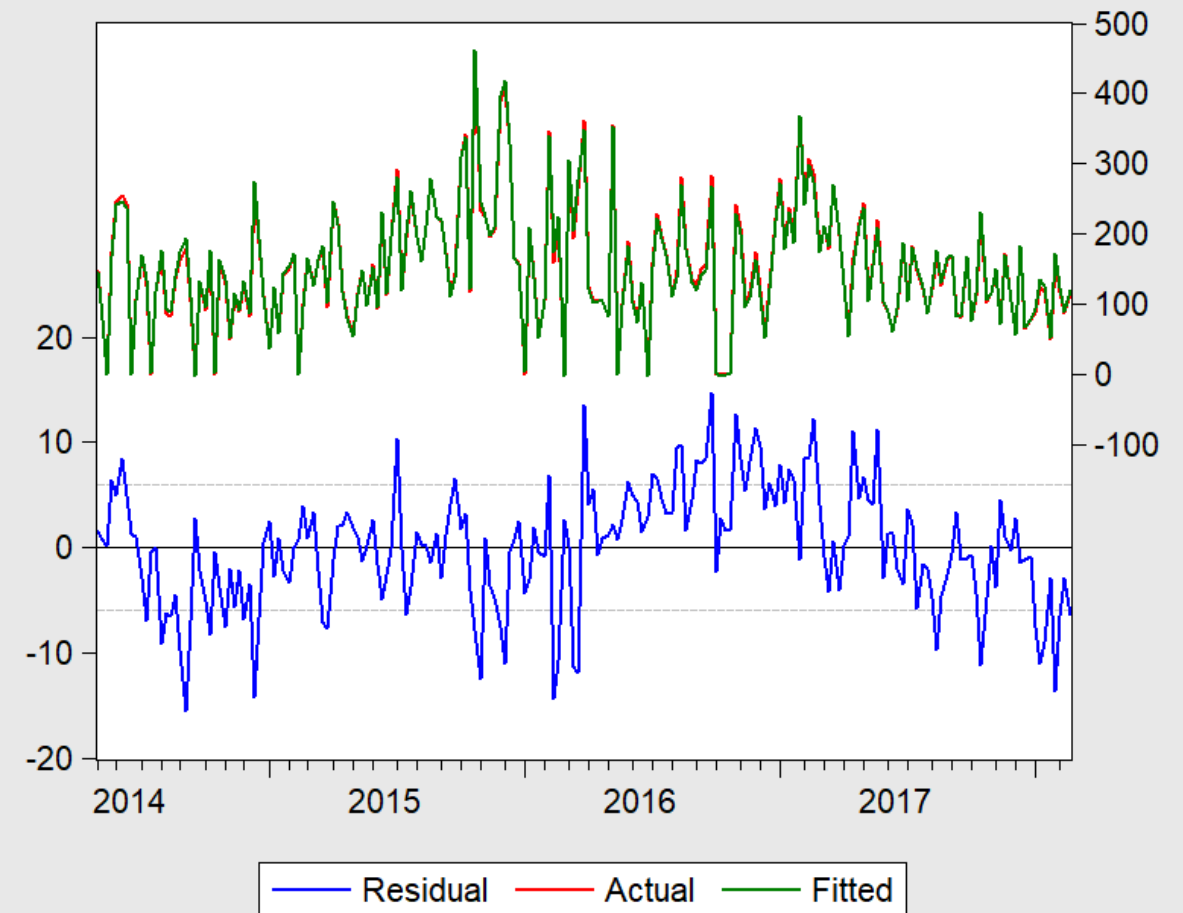
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-4.324936	2.323115	-1.861697	0.0642
C(2)	0.018587	0.007219	2.574868	0.0108
C(3)	0.004040	0.005511	0.733181	0.4644
C(4)	0.008916	0.005527	1.613206	0.1084
C(5)	0.006057	0.005250	1.153602	0.2501
C(6)	0.003014	0.004997	0.603109	0.5472
C(7)	-0.001105	0.000194	-5.688547	0.0000
C(8)	3762.807	664.5075	5.662550	0.0000
C(9)	0.241458	0.084632	2.853025	0.0048
C(10)	-14.76087	5.618582	-2.627153	0.0093
C(11)	-89.02550	257.9287	-0.345155	0.7304
C(12)	641.5162	5.401003	118.7772	0.0000

R-squared	0.995449	Mean dependent var	152.2398
Adjusted R-squared	0.995181	S.D. dependent var	85.40468
S.E. of regression	5.928488	Akaike info criterion	6.455822
Sum squared resid	6572.484	Schwarz criterion	6.654414
Log likelihood	-630.3543	Hannan-Quinn criter.	6.536197
F-statistic	3718.493	Durbin-Watson stat	0.915864
Prob(F-statistic)	0.000000		

Reach

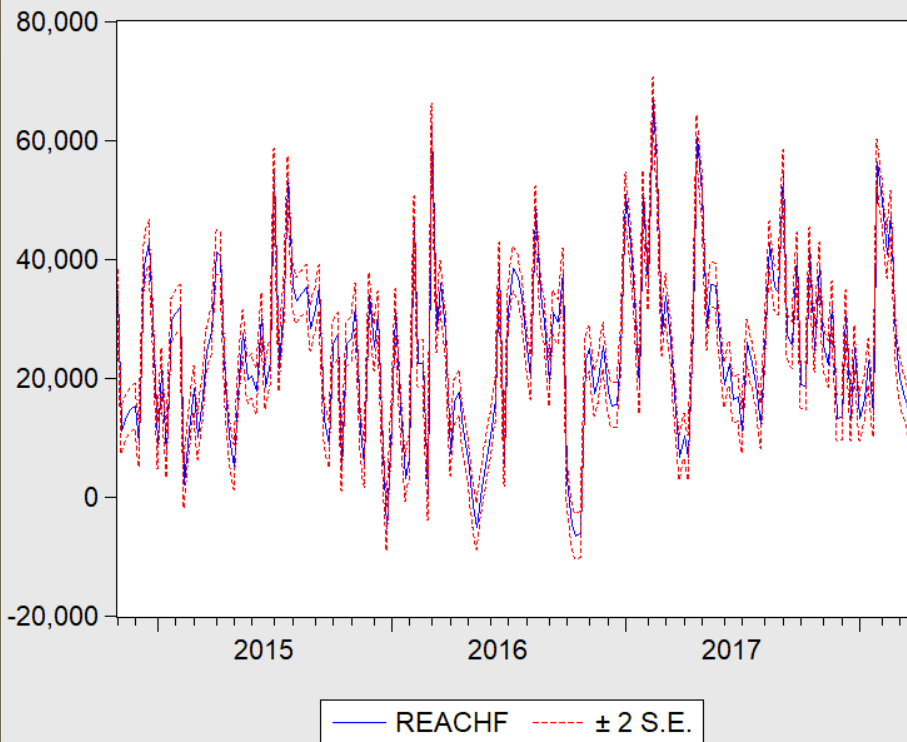


Affinity

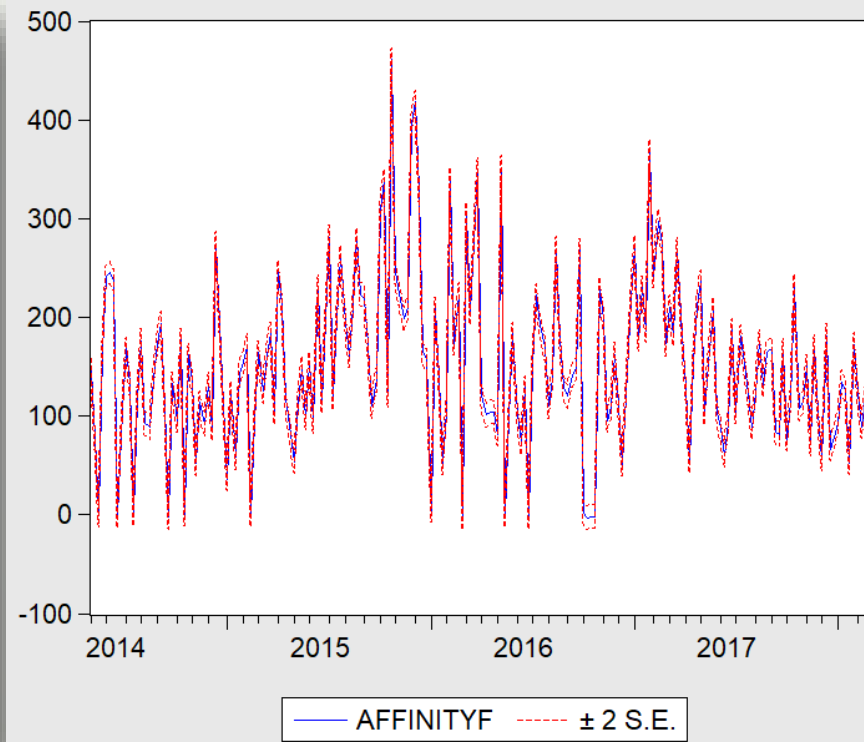


Результати роботи у системі EViews. Прогнозування

Reach



Affinity



Forecast: AFFINITYF
Actual: AFFINITY
Forecast sample: 12/30/2013 3/26/2018
Adjusted sample: 5/05/2014 3/19/2018
Included observations: 203

Root Mean Squared Error	5.802790
Mean Absolute Error	4.496940
Mean Abs. Percent Error	2.967732
Theil Inequality Coefficient	0.016786
Bias Proportion	0.000083
Variance Proportion	0.002126
Covariance Proportion	0.997791

Результати роботи власної програми

Прогнозування трафіку в сфері Інтернет-реклами

Вибір сайту для прогнозування

Шлях до даних

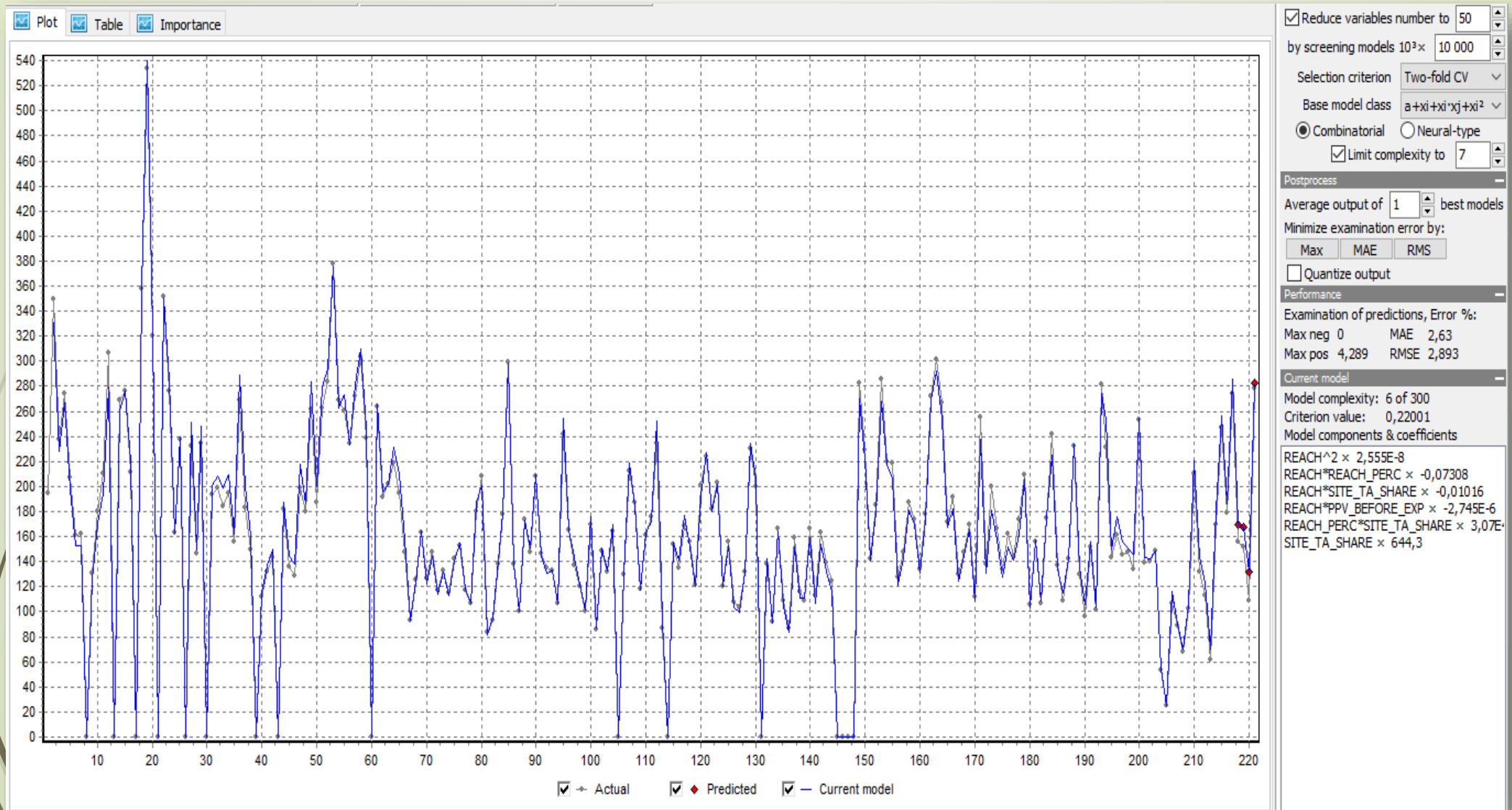
Назва сайту

Категорія сайту

Кількість кроків прогнозу

	Назва	Категорія
	isport.ua	Спортивна
▶	football24.ua	Спортивна
*		

	Крок прогнозу	Назва сайту	Категорія сайту	Охоплення	Аффініті індекс
	1	isport.ua	Спортивна	23429.28345	64.81295
	2	isport.ua	Спортивна	21319.24931	60.49101
	3	isport.ua	Спортивна	19129.94882	37.62387
	4	isport.ua	Спортивна	16681.57645	55.53669
	1	football24.ua	Спортивна	43350.04975	163.84678
▶	2	football24.ua	Спортивна	27698.71734	156.12476



Порівняння результатів

Метод	Назва сайту	Прогнозований медіапоказник	MAPE	Коментарі
МГУА	isport.ua	reach	9,23%	Комбінаторний алгоритм, складність 7
	isport.ua	affinity	11,31%	Комбінаторний алгоритм, складність 5
МГУА	football24.ua	reach	3,85%	Комбінаторний алгоритм, складність 7
	football24.ua	affinity	10,30%	Комбінаторний алгоритм, складність 7
АР	isport.ua	reach	13,49%	Динамічний прогноз
	isport.ua	affinity	2,96%	Динамічний прогноз
АР	football24.ua	reach	16,48%	Динамічний прогноз
	football24.ua	affinity	2,71%	Динамічний прогноз
Власна програма	isport.ua	reach	13,49%	Програмувався авторегресійний метод
	isport.ua	affinity	2,96%	Програмувався авторегресійний метод
Власна програма	football24.ua	reach	16,48%	Програмувався авторегресійний метод
	football24.ua	affinity	2,71%	Програмувався авторегресійний метод

Порівняння результатів

	Реальні дані	AP	МГУА
isport.ua (Reach)	23428,00	23429,29	21490,60
	23309,00	21319,25	21367,10
	19223,00	19129,95	17156,60
	17368,00	16681,58	15707,10
football24.ua (Reach)	44765,00	43350,05	42671,70
	25849,00	27698,72	23905,70
	77257,00	72992,16	72202,60
	80727,00	71965,73	83408,30
isport.ua (Affinity)	63,10	64,81	70,31
	54,73	60,49	61,07
	35,86	37,63	39,76
	53,51	55,54	59,58
football24.ua (Affinity)	156,39	163,85	169,44
	152,11	156,13	167,46
	108,83	120,92	131,70
	278,15	297,76	282,99

Висновки та рекомендації щодо подальшої роботи

- За допомогою авторегресійного методу та методу групового урахування аргументів вдалося отримати високоякісні прогнози найважливіших для нас змінних.
- Обидва методи спрацювали найкращим чином, тому кожен з них можна використовувати для вирішення даної або подібній даній задач.
- Був створений власний програмний продукт з метою прискорення та полегшення процесу прогнозування. Інтерфейс зроблено максимально зручним і інтуїтивно зрозумілим у користуванні.

В подальшому результати прогнозування можна покращити, якщо будуть реєструватись додаткові дані.

Результати роботи можна використовувати для модифікації вже існуючих програмних забезпечень.

Дякую за увагу!