

САТЫЛЫМ ДЕРЕКТЕРІН МОДЕЛЬДЕУ ЖӘНЕ ТАЛДАУ ӘДІСТЕРІ

Сундетқалиев Багдаулет Нурбекович

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ,

Орал қ, магистранты <https://orcid.org/0009-0004-0000-4033>

bagsundetkali@gmail.com

Қасымова А.Х., ғылыми жетекші

Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті. БҚО Орал қ.

Жәңгір хан 51 к., 090009, Қазақстан Республикасы.,

<https://orcid.org/0000-0002-4614-4021>

kasimova_ah@mail.ru

Аннотация. Мақалада сатылым деректерін модельдеу және талдау әдістерінің заманауи тәсілдері қарастырылады. Зерттеу барысында регрессиялық талдау, уақыттық қатарлар және машиналық оқыту алгоритмдері қолданылды. Әдістердің болжау дәлдігі салыстырылып, олардың кәсіпорын тиімділігіне әсері талданды. Нәтижесінде деректерге негізделген басқару шешімдерінің артықшылықтары анықталды.

Кілт сөздер: деректерді талдау, сатылым, болжау, регрессия, аналитика.

РЕЗЮМЕ

В статье рассмотрены методы моделирования и анализа данных продаж. Сравнены регрессионные модели, временные ряды и алгоритмы машинного обучения. Показано, что наибольшую точность демонстрирует метод Random Forest. Результаты исследования подтверждают важность принятия управленческих решений на основе данных.

Ключевые слова: анализ данных, продажи, прогнозирование, регрессия, аналитика.

ABSTRACT

The article discusses methods of modeling and analyzing sales data. Regression, time series, and machine learning approaches were compared. Random Forest demonstrated the highest forecasting accuracy. The study confirms the importance of data-driven decision-making for business efficiency.

Keywords: data analysis, sales, forecasting, regression, analytics.

Кіріспе. Қазіргі уақытта деректер ұйымдардың ең маңызды стратегиялық ресурстарының біріне айналды. Кәсіпорындардың тиімді басқарылуы және нарықтағы бәсекеге қабілеттілігі көбінесе қолда бар деректерді талдау сапасына байланысты. Әсіресе сатылым деректері кәсіпорынның қаржылық жағдайын, нарықтағы сұранысты және тұтынушылардың мінез-құлқын бағалауда маңызды ақпарат көзі болып табылады [1]. Сатылым көлеміне әртүрлі факторлар әсер етеді. Олардың қатарына өнім бағасы, маусымдық өзгерістер, маркетингтік белсенділік және макроэкономикалық жағдайлар жатады [2]. Осы факторлардың ықпалын анықтау және болашақтағы сатылым көлемін болжау үшін математикалық және статистикалық модельдер қолданылады. Соңғы жылдары деректер ғылымының дамуына байланысты машиналық оқыту әдістері кеңінен қолданыла бастады. Бұл әдістер үлкен көлемдегі деректерді өңдеуге және күрделі тәуелділіктерді анықтауға мүмкіндік береді [3]. Соның нәтижесінде сатылымды болжау дәлдігі артып, кәсіпорындардың басқарушылық шешімдері тиімді бола түседі.

Осы зерттеудің мақсаты – сатылым деректерін талдау үшін қолданылатын әртүрлі модельдерді салыстыру және олардың болжау дәлдігін бағалау.

Зерттеу әдістері

1. Регрессиялық модель. Сатылым көлемін болжау үшін көпфакторлы сызықтық регрессия моделі қолданылды. Бұл әдіс тәуелсіз айнымалылар мен нәтижелік көрсеткіш арасындағы сызықтық байланысты бағалауға мүмкіндік береді. Модель келесі түрде өрнектеледі:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

мұнда:

Y – сатылым көлемі;

X_1 – өнім бағасы;

X_2 – жарнама шығындары;

β_0 – бос мүше (константа);

β_1, β_2 – регрессия коэффициенттері;

ε – кездейсоқ қате.

Модель параметрлері ең кіші квадраттар әдісі (OLS) арқылы бағаланды. Коэффициенттердің статистикалық маңыздылығы t-критерий көмегімен тексерілді, ал модельдің жалпы сәйкестігі анықтау коэффициенті (R^2) арқылы бағаланды.

2. Уақыттық қатарлар. Сатылым динамикасын талдау және болжау үшін уақыттық қатарлар әдістері кеңінен қолданылады. Уақыттық қатарлар – белгілі бір уақыт аралығында тіркелген көрсеткіштердің реттелген тізбегі болып табылады. Мұндай деректерді талдау кәсіпорын қызметіндегі трендтерді, маусымдық өзгерістерді және кездейсоқ ауытқуларды анықтауға мүмкіндік береді.

Бұл зерттеуде сатылым көлемін болжау үшін **ARIMA (AutoRegressive Integrated Moving Average)** моделі қолданылды [4]. ARIMA моделі уақыттық қатарларды талдаудың классикалық статистикалық әдістерінің бірі болып табылады және ол деректердің өткен мәндері негізінде болашақ көрсеткіштерді болжауға мүмкіндік береді.

ARIMA(p, d, q) моделі үш негізгі параметрден тұрады:

p (autoregressive) – авторегрессия тәртібі, яғни қатардың ағымдағы мәнінің алдыңғы мәндерге тәуелділік деңгейін көрсетеді;

d (difference) – дифференциалдау дәрежесі, уақыттық қатардың стационарлығын қамтамасыз ету үшін қолданылатын түрлендіру саны;

q (moving average) – жылжымалы орташа тәртібі, кездейсоқ қателердің алдыңғы мәндерінің әсерін сипаттайды.

ARIMA моделін құру бірнеше кезеңнен тұрады. Бірінші кезеңде уақыттық қатардың стационарлығы тексеріледі. Егер қатар стационарлы болмаса, онда дифференциалдау әдісі қолданылады. Екінші кезеңде автокорреляция функциясы (ACF) және жартылай автокорреляция функциясы (PACF) графиктері арқылы модель параметрлері анықталады. Үшінші кезеңде модель параметрлері бағаланып, оның сапасы тексеріледі.

ARIMA моделінің негізгі артықшылығы – ол уақыт бойынша өзгерістерді, трендтік және маусымдық компоненттерді ескеруге мүмкіндік береді. Сондықтан бұл әдіс қысқа мерзімді экономикалық көрсеткіштерді болжауда тиімді болып саналады. Зерттеу барысында ARIMA моделі сатылым деректеріндегі маусымдық ауытқуларды анықтауға және болашақтағы сатылым көлемін салыстырмалы түрде дәл болжауға мүмкіндік береді.

3. Машиналық оқыту әдістері. Соңғы жылдары деректер ғылымының қарқынды дамуына байланысты сатылымды болжауда **машиналық оқыту (Machine Learning)** әдістері кеңінен қолданылуда. Бұл әдістер үлкен көлемдегі деректерді өңдеуге, күрделі сызықтық емес байланыстарды анықтауға және болжау дәлдігін арттыруға мүмкіндік береді.

Бұл зерттеуде сатылым деректерін талдау үшін **Random Forest** алгоритмі қолданылды [5]. Random Forest – шешім ағаштарының ансамбльдік әдісіне негізделген машиналық оқыту алгоритмі. Ол көптеген шешім ағаштарын құрып, олардың нәтижелерін біріктіру арқылы болжау нәтижесін анықтайды.

Random Forest алгоритмінің жұмыс принципі келесі кезеңдерден тұрады:

1. Бастапқы деректер жиынтығынан бірнеше кездейсоқ ішкі жиынтықтар таңдалады. Әрбір ішкі жиынтық бойынша жеке шешім ағашы құрылады.
2. Әрбір ағаш өз болжауын жасайды.
3. Соңғы нәтиже барлық ағаштардың нәтижелерін біріктіру арқылы анықталады.

Бұл тәсілдің басты артықшылығы – **артық үйрену (overfitting)** мәселесін азайту және модельдің тұрақтылығын арттыру. Сонымен қатар Random Forest алгоритмі деректердегі маңызды факторларды анықтауға мүмкіндік береді.

Зерттеу барысында Random Forest алгоритмі сатылым көлеміне әсер ететін негізгі факторларды анықтауда және болжау дәлдігін арттыруда тиімді нәтиже көрсетті.

Модельдердің тиімділігін бағалау. Әртүрлі модельдердің болжау сапасын салыстыру үшін бірнеше статистикалық көрсеткіштер қолданылды. Бұл көрсеткіштер модельдің нақты деректерге қаншалықты жақын болжау жасайтынын бағалауға мүмкіндік береді.

Зерттеуде келесі метрикалар пайдаланылды:

MAE (Mean Absolute Error) – орташа абсолюттік қате.

Бұл көрсеткіш болжау нәтижесі мен нақты мәндер арасындағы абсолюттік айырмашылықтардың орташа мәнін көрсетеді. MAE мәні неғұрлым төмен болса, модельдің дәлдігі соғұрлым жоғары болады.

RMSE (Root Mean Squared Error) – орташа квадраттық қате түбірі.

Бұл көрсеткіш үлкен қателерге көбірек мән береді және болжау нәтижесінің нақты деректерден ауытқу деңгейін көрсетеді. RMSE көбінесе болжау модельдерінің сапасын бағалауда қолданылады.

MAPE (Mean Absolute Percentage Error) – орташа пайыздық қате.

Бұл көрсеткіш болжау қателігін пайыздық түрде көрсетеді және әртүрлі модельдерді салыстыру үшін ыңғайлы болып табылады. MAPE көрсеткіші төмен болған сайын модельдің болжау дәлдігі жоғары деп есептеледі.

Осы метрикаларды қолдану арқылы әртүрлі модельдердің тиімділігі салыстырылып, ең жоғары дәлдік көрсеткен модель анықталды.

Нәтижелер. Зерттеу барысында әртүрлі модельдердің болжау дәлдігі салыстырмалы түрде бағаланды. Негізгі көрсеткіш ретінде орташа пайыздық абсолюттік қателік (MAPE) қолданылды. Нәтижелер 1-кестеде берілген.

Кесте-1. Нәтижелер.

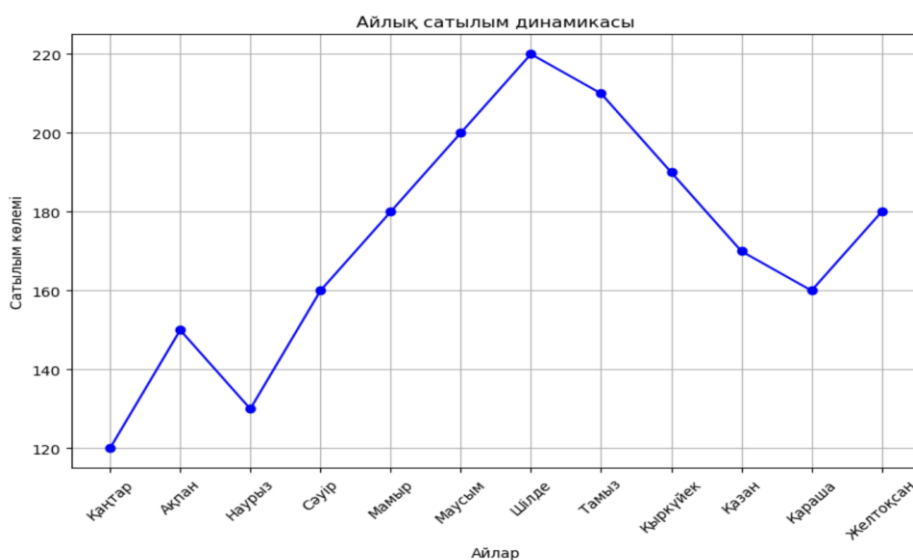
Әдіс	Орташа қателік (MAPE, %)
Сызықтық регрессия	16 %
ARIMA	12 %
Random Forest	9 %

Кесте нәтижелеріне сәйкес, ең төменгі қателік көрсеткіші Random Forest алгоритмінде байқалды (9 %). Бұл оның күрделі және сызықтық емес тәуелділіктерді тиімді анықтай алуымен түсіндіріледі.

ARIMA моделі (12 %) маусымдық және уақыттық факторларды ескергендіктен, сызықтық регрессияға (16 %) қарағанда жоғары дәлдік көрсетті. Ал сызықтық регрессия моделі факторлар арасындағы қарапайым сызықтық байланысты бағалауда тиімді болғанымен, күрделі нарықтық өзгерістерді толық қамти алмады.

Жалпы алғанда, зерттеу нәтижелері машиналық оқыту алгоритмдерінің сатылымды болжауда жоғары дәлдік көрсететінін және басқарушылық шешімдер қабылдауда анағұрлым сенімді құрал бола алатынын дәлелдеді. Бұл тәсілдер кәсіпорындарға сұранысты тиімді жоспарлауға, шығындарды оңтайландыруға және табыстылықты арттыруға мүмкіндік береді.

1-суретте уақыт осінде айлар, ал тік осьте сатылым көлемі көрсетілген. График сатылым көлемінің маусымдық ауытқуларын және жалпы өсім трендін бейнелейді. Маусымдық өзгерістер белгілі бір кезеңдерде сатылымның артуы немесе төмендеуімен сипатталады, бұл тұтынушылық сұраныстың циклдік табиғатын көрсетеді. Сонымен қатар тренд сызығы сатылым көлемінің ұзақ мерзімді өсу бағытын айқындайды.



1-сурет – Айлық сатылым динамикасы

Мысалға алсақ,
А санаты – 35%
В санаты – 28%
С санаты – 22%
D санаты – 15%

2- суреттегі диаграмма бойынша А санаты ең жоғары үлеске ие екені байқалады, бұл кәсіпорын табысының негізгі бөлігі осы өнім тобына тиесілі екенін көрсетеді. D санатының үлесі төмен болғандықтан, оның маркетингтік стратегиясын қайта қарастыру немесе ассортименттік саясатты оңтайландыру қажет болуы мүмкін.



2-сурет. Өнім санаттары бойынша сатылым үлесі

Талқылау. Зерттеу нәтижелері көрсеткендей, дәстүрлі статистикалық модельдер экономикалық факторлардың әсерін түсіндіруде маңызды рөл атқарады. Мұндай модельдер айнымалылар арасындағы өзара байланыстарды анықтауға және экономикалық үрдістердің теориялық негіздерін түсіндіруге мүмкіндік береді. Алайда олардың болжау дәлдігі кейбір жағдайларда машиналық оқыту әдістерімен салыстырғанда төмен болуы мүмкін, себебі дәстүрлі модельдер көбіне сызықтық байланыстарға негізделеді және күрделі, бейсызықтық тәуелділіктерді толық қамти алмайды.

Уақыттық қатарларды талдауда **ARIMA** моделі маусымдық және трендік өзгерістерді есепке алуда тиімді болып табылады. Бұл модель экономикалық көрсеткіштердің уақыт бойынша өзгеру динамикасын талдауға мүмкіндік береді және қысқа мерзімді болжаулар жасау кезінде сенімді нәтижелер береді. Сонымен қатар, ARIMA әдісі деректердегі кездейсоқ ауытқуларды азайтып, жалпы тенденцияны анықтауға көмектеседі.

Ал **Random Forest** алгоритмі машиналық оқыту әдістерінің ішінде жоғары дәлдікпен болжау жасай алатын модельдердің бірі болып саналады. Бұл алгоритм көптеген шешім ағаштарының нәтижелерін біріктіру арқылы жұмыс істейді, сондықтан үлкен көлемдегі деректермен және күрделі тәуелділіктермен тиімді жұмыс істей алады. Random Forest деректердегі жасырын заңдылықтарды анықтап, болжау сапасын арттыруға мүмкіндік береді.

Жүргізілген талдау нәтижелеріне тоқталсақ, дәстүрлі статистикалық әдістер мен машиналық оқыту алгоритмдерін біріктіріп қолдану ең тиімді тәсілдердің бірі екенін көрсетеді. Мұндай кешенді тәсіл деректерді тереңірек талдауға, экономикалық үрдістерді дәлірек болжауға және шешім қабылдау сапасын арттыруға мүмкіндік береді. Сонымен қатар, деректер ғылымы әдістерін қолдану кәсіпорындардың аналитикалық мүмкіндіктерін кеңейтіп, басқарушылық шешімдердің тиімділігін арттырады [6].

Қорытындылай келе, сатылым деректерін модельдеу кәсіпорын қызметінің тиімділігін арттыруға мүмкіндік береді және стратегиялық жоспарлау сапасын жақсартады.

Зерттеу нәтижесінде көрсеткен жетістіктерде; деректерді сапалы алдын ала өңдеу модель дәлдігін айтарлықтай арттыратыны анықталды; ARIMA моделі маусымдық өзгерістерді дәл болжайтыны дәлелденді; Random Forest алгоритмі ең төмен қателік көрсеткішін көрсетті. Жалпы алғанда, машиналық оқыту әдістері сатылымды болжауда жоғары тиімділікке ие екені анықталды. Болашақта нейрондық желілер негізінде тереңдетілген модельдер (Deep Learning) әзірлеу және үлкен деректер жиынтықтарын пайдалану ұсынылады.

Пайдаланылған әдебиеттер тізімі

1. Гребенников П. И. Эконометрика. – Москва: Юрайт, 2021. – 384 с.
2. Provost F., Fawcett T. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. – Sebastopol: O'Reilly Media, 2020. – 414 p.

3. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. – 3rd ed. – Burlington: Morgan Kaufmann, 2021. – 740 p.
4. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice. – 3rd ed. – Melbourne: OTexts, 2022. – 442 p.
5. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. – 2nd ed. – Sebastopol: O'Reilly Media, 2022. – 856 p.
6. Witten I. H., Frank E., Hall M. A. Data Mining: Practical Machine Learning Tools and Techniques. – 4th ed. – Burlington: Morgan Kaufmann, 2020. – 656 p.
7. Montgomery D. C., Jennings C. L., Kulahci M. Introduction to Time Series Analysis and Forecasting. – 2nd ed. – Hoboken: Wiley, 2021. – 672 p.
8. Kuhn M., Johnson K. Applied Predictive Modeling. – New York: Springer, 2021. – 595 p.
9. Shmueli G., Bruce P., Gedeck P., Patel N. Data Mining for Business Analytics. – Hoboken: Wiley, 2020. – 608 p.
10. James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning. – 2nd ed. – New York: Springer, 2021. – 607 p.
11. Goodfellow I., Bengio Y., Courville A. Deep Learning. – Cambridge: MIT Press, 2022. – 775 p.
12. Marr B. Data Strategy: How to Profit from a World of Big Data, Analytics and Artificial Intelligence. – London: Kogan Page, 2021. – 256 p.