

## Моделі прогнозування автомобільного трафіку на основі LSTM при розробці прикладних програм

У роботі досліджується проблема прогнозування автомобільного трафіку, яка є важливою для мешканців сучасних міст і їх груп при плануванні переміщень у місті зокрема і власних справ в цілому, враховуючи підвищення забрудненості повітря зі збільшенням трафіку. Мета роботи полягала в дослідженні впливу різних входних ознак LSTM-моделі та тривалості входного часового інтервалу на результати прогнозування автомобільного трафіку в місті. Запропоновані рішення дозволяють подолати обмеження щодо необхідності наявності цілої системи станцій спостереження та великого набору історичних даних, які виникають при використанні моделей, запропонованих у дослідженнях інших авторів. У роботі задача прогнозування автомобільного трафіку, яка є задачею прогнозування часового ряду, розглянута у перспективі 6 наступних годин з інтеграцією створених моделей до прикладних програм, для чого визначено відповідні процедури збору та оброблення даних. Проведено експериментальне дослідження на основі даних, зібраних за 59 станціями спостереження, оцінювання результатів якого відбувається за рядом показників (середньої абсолютної, середньоквадратичної і кореня середньоквадратичної похибки) щодо точності та за коефіцієнтом детермінації стосовно інформативності. Запропоновані моделі прогнозування побудовані на основі використання довгої короткочасної пам'яті. Експериментальне дослідження підтвердило підвищення точності та інформативності створених моделей при використанні більш тривалого входного часового інтервалу (24 годин замість 6) та додаткових входних ознак на основі даних трафіку станцій спостереження, які були відібрані зі всього набору станцій за допомогою ансамблю дерев рішень на основі методу Random Forest.

**Ключові слова:** автомобільний трафік; прогнозування; часовий ряд; машинне навчання; довга короткочасна пам'ять.

**Актуальність теми.** Автомобільний трафік впливає на життя будь-якої людини в місті, не зважаючи на те, чи вона є водієм автомобіля, чи лише користується громадським транспортом. До того ж керування автомобільним трафіком є відповідальністю осіб, які ухвалюють управлінські рішення в місті. Тому розв'язання проблем, пов'язаних з автомобільним трафіком, впливає на життя будь-якої людини в місті. Як відомо, кількість людей у світі, які проживають у містах, збільшується. Зокрема за даними ООН на даний момент у світі більше ніж 50 % людей проживають у містах, а за прогнозами до 2050 року ця частка досягне 68 %, що буде дорівнювати 6,7 мільярдам людей [1].

У країнах, які розвиваються, кількість власного транспорту мешканців збільшується. Це призводить до ускладнення переміщень у місті, впливає на вибір транспортних засобів мешканцями. Зокрема в європейських країнах донедавна активно використовувалися малогабаритні автомобілі. У деяких містах популярними є мотоцикли, мопеди або інші подібні типи транспорту. Все це фактично є відображенням того, який рівень трафіку в місті, як організовано транспортні потоки, і відповідно відображає складність цих процесів у системі керування містом. Без комплексного керування всіма процесами в сучасному місті досягти зручності життя мешканців, відвідувачів неможливо.

Проте проблеми, пов'язані з автомобільним трафіком, є актуальними не лише в стратегічному вимірі, але і в короткотривалій перспективі. Прогнозування автомобільного трафіку на такий період є важливим для індивідуальних мешканців, груп мешканців, а також і для відповідного керування транспортними потоками. У першому випадку мешканці можуть спланувати власні поїздки таким чином, щоб витратити менше зайвого часу, адже проблема тривалого переміщення з однієї точки в місті в іншу через ускладнений трафік нерідко посилюється за використання власного транспорту тим, що додатковий час витрачається на пошук місця для паркування. Водночас через пошук місця для паркування власниками автомобілів трафік збільшується додатково. Тож отримання інформації про завантаженість потоків автомобільного трафіку також може дозволити автовласникам вчасно уникати тих місць, де відбуваються тимчасові скупчення. При керуванні містом прогнозування на такі періоди дозволяє швидко реагувати на те, як певні події, обмеження впливають фактично на наявні потоки. Тож в цілому все це вказує на актуальність проблеми прогнозування автомобільного трафіку в місті.

**Аналіз останніх досліджень та публікацій, на які спирається автор.** На даний момент можна виділити ряд робіт, які стосуються прогнозування автомобільного трафіку.

У роботах [2–3] опубліковані результати досліджень, що ґрунтуються на застосуванні складних гібридних моделей на основі згорткових нейронних мереж, довгої короткочасної пам'яті або long short-term memory (LSTM), графових нейронних мереж, що для виділення взаємного розташування станцій спостереження за трафіком потребує наявності таких станцій на всіх або великій кількості доріг міста. Тоді це дозволяє моделі робити висновки про те, як одні автомобільні потоки впливають на інші. Однак не завжди необхідні дані для роботи таких моделей наявні: кількість станцій спостереження може бути невеликою (навіть не зважаючи на наявність оригінальних пропозицій як, наприклад, у роботі [4], організувати постійний збір даних за кожною автодорогою достатньо складно і витратно), ряд історичних даних може бути значно обмеженим.

У роботі [5] прогнозування трафіку виконується на 5 хвилин за минулими 30 хвилинами за допомогою моделей AutoRegressive Integrated Moving Average (ARIMA), LSTM і Gated Recurrent Unit (GRU), які призвели до отримання кращих результатів. Дослідження [6] використовувало метод опорних векторів, радіально-базисні нейронні мережі, ARIMA, стековий автокодувальник, рекурентну нейронну мережу та LSTM, остання з яких при прогнозуванні на 15, 30, 45, 60 хвилин дозволила отримати найкращі результати. Використання LSTM-моделей виконується також у роботі [7], а в роботі [8] застосовуються двонаправлені LSTM (BiLSTM) для прогнозування на період у 5, 10, 15, 30, 45, 60 хвилин за допомогою модельованих даних.

В цілому роботи [5–8] демонструють загальну практичну придатність LSTM-моделей для розв'язання задачі прогнозування автомобільного трафіку. Існують варіанти, що враховують різні особливості структури моделей, проте в цілому дані роботи вказують на переваги моделей на основі LSTM відносно інших моделей та методів машинного навчання. Проте, якщо перша група методів, яку важко застосувати за обмеженості даних, використовує в якості вхідних ознак дані всіх станцій, то моделі в роботах [5–8] фактично використовують історичні дані для станції, за якою будується модель. А для кожної станції при цьому створюються окремі моделі, на відміну від логіки роботи першої групи рішень. З іншого боку при створенні моделей для окремих станцій додатковими вхідними ознаками розглядаються лише деякі метеорологічні дані. Однак можливий вплив даних інших станцій, збільшення часового інтервалу вхідних даних у роботах не досліджується.

**Метою статті** є дослідження впливу різних вхідних ознак LSTM-моделі та тривалості вхідного часового інтервалу на результати прогнозування автомобільного трафіку в місті.

**Викладення основного матеріалу.** Створення і подальше застосування LSTM-моделей [9] для прогнозування автомобільного трафіку потребує визначення цілого ряду параметрів, частина з яких є більш вивченою, а частина – ні. Тому в даному дослідженні необхідно було вивчити те, як впливає збільшення періоду історичних даних на точність прогнозування. Зокрема це стосується того, чи призводить до збільшення точності прогнозування збільшення величини вхідного часового інтервалу до значення, більшого ніж величина часового інтервалу, на який відбувається прогнозування автомобільного трафіку. Існуючі роботи зазвичай використовують тривалість вхідного часового інтервалу, рівну тривалості вихідного. Проте і прогнозування в них найчастіше відбувається на короткий період часу, який виражається в хвилинах, але триває до години.

У межах даного дослідження необхідно було дослідити прогнозування на більш тривалий період, при чому в якості бази потрібно використовувати період в годину. Тобто кожен екземпляр даних, який обробляється моделлю, є часовим рядом, кожна точка якого представляє кількість автотранспорту, який подолав станцію вимірювання протягом години. Такий підхід призводить зокрема до того, що період історичних спостережень, протягом якого має бути сформована вибірка даних, значно збільшується. Тобто якщо використовувати дані за кожну хвилину, то за використання години період спостереження має збільшитись у 60 разів. Якщо використовуються накопичені дані за кожні 5 хвилин, то відповідно в 12. Такі обмеження виникають через те, що модель певної структури для свого якісного навчання потребує певної кількості екземплярів даних, яка має бути не меншою ніж деякий поріг, який визначається, виходячи з кількості параметрів моделі. Тож у підсумку, якщо за використання щохвилинних даних потрібно накопичити дані за місяць для навчання деякої моделі, то за використання щогодинних даних для моделі тієї ж самої структури знадобиться 5 років. Вимоги до даних зростають.

Окрім того вивчені моделі найчастіше на вхід отримують історичні дані стосовно тієї станції, для якої модель створюється. Проте можливо інформативними даними є дані, які характеризують трафік за іншими станціями. Збільшення трафіку на одній станції в певний момент може призвести до збільшення трафіку на іншій станції протягом наступної години. Тобто якщо виникає затор, розпочинається певна подія тощо, то такі дані можуть бути важливими для кожної конкретної моделі.

Тому в підсумку було вирішено створювати LSTM-моделі, які відрізняються кількістю вхідних ознак та тривалістю вхідного часового інтервалу для цих ознак.

Задача прогнозування автомобільного трафіку є задачею прогнозування часового ряду. Тобто на вхід моделі, яка має бути створена для розв'язання задачі, подаються дані, які характеризують  $P$  послідовних годин у минулому, а на виході модель має призвести до отримання значень трафіку за наступні  $F$  годин,

тобто за  $P + 1, P + 2, \dots, P + F$  години. У процесі дослідження мають бути визначені ознаки, за якими на вхід моделі мають подаватися відповідні послідовні  $P$  значень.

Для визначення впливу ознак, параметрів створювалися окремі моделі, які представляли окремі варіанти підмножин вхідних ознак, окремі значення параметрів. Але для підтримки навчання та тестування всіх цих моделей потрібно було спочатку створити вибірку даних про автомобільний трафік. Це фактично і є особливістю інтеграції моделей прогнозування автомобільного трафіку під час розробки прикладних програм. Такі прикладні програми можуть використовуватися:

- безпосередньо для прогнозування автомобільного трафіку;
- визначення оптимальних маршрутів переміщення між локаціями в місті;
- аналізу потоків забруднювачів повітря в місті;
- моніторингу стану трафіку зокрема та процесів у місті в цілому тощо.

У випадку створення такого програмного забезпечення відповідно робота відбувається не за статичною сформованою заздалегідь вибіркою даних. Моделі мають оновлюватися в процесі роботи, використовуючи для навчання нові отримані дані. Відповідно ці дані мають бути завантажені спочатку з ресурсів, які щонайменше нерідко є зовнішніми, потім оброблені необхідним чином дозволяють сформувати вибірку, яка вже використовується для роботи з моделлю. Окрім оновлення моделі також нові отримані дані постійно мають використовуватися для виконання нових обчислень, а для того, щоб подати їх на вхід моделі, спочатку ці дані потрібно обробити. Тому процес оброблення даних стає важливим початковим процесом, який має бути забезпечений для формування вибірки даних і який має підтримуватися в подальшому на кожному кроці для роботи. Наприклад, якщо прогнозування відбувається на деяку кількість годин у майбутньому, то ці дії повторюються кожен годину, відповідно і оброблення даних має реалізовуватися кожен годину над всіма новими екземплярами.

Для виконання оброблення даних спочатку мають бути завантажені дані. Для цього або використовуються посилання на відповідні файли, розташовані на сервері, імена яких формуються згідно з необхідним періодом, або використовується Application Programming Interface (API), якщо він надається для отримання відповідних даних.

Перед отриманням даних має бути визначено період навчальної вибірки. За кожною мінімальною датою даного періоду (година, день, місяць, в залежності від того, якими порціями можуть бути завантажені дані) має бути встановлено, чи наявні ці дані у завантажених і збережених даних у базі даних. Якщо вони наявні, то дані не завантажуються. Якщо ні, то запитуються з сервера. Після отримання даних вони мають бути розподілені за станціями та зведені за кожною станцією до кожної часової відмітки, що відповідає встановленому періоду навчання. Якщо мінімальною одиницею є година, то відповідно всі години в періоді навчання мають бути визначені та внесені у вибірку за кожною відповідною станцією. А за кожною годиною, яка визначає відповідний екземпляр даних, має бути встановлено значення трафіку за цю годину за цією станцією. За використання додаткових даних (наприклад, метеорологічних), які співвідносяться зі станціями спостереження за трафіком, вони мають бути додані відповідними значеннями додаткових ознак за кожною годиною. Зокрема у випадку метеорологічних даних, наприклад, для кожної станції спостереження за трафіком може бути встановлена найближча метеорологічна станція за кожним необхідним показником.

Далі отримані дані мають бути проаналізовані щодо виявлення некоректних значень. Це значення, які мають несумісний тип, не потрапляють у відведений для показника діапазон значень, або пропущені значення. Усі такі значення тоді фактично визначаються як пропущені та виконується лінійна інтерполяція. Необхідність цих дій продиктована тим, що для роботи моделі має бути подано на вхід не окреме значення, а вся послідовність (вхідна частина часового ряду). Тому якщо певне значення пропущене (наприклад, встановлене як нульове), то модель буде працювати некоректно. Після цього має бути виконано нормування за кожним показником. Нормування реалізовувалося, визначаючи мінімальне та максимальне значення для кожного окремого показника (трафік, метеорологічні дані) для кожної окремої станції, а тоді зводячи всі значення за даним показником до інтервалу від 0 до 1. Оскільки під час експериментального дослідження не всі дані використовувались для навчання, то знаходження мінімального та максимального значень виконувалося лише за навчальною вибіркою.

Для проведення експериментального дослідження було використано дані спостережень за трафіком у місті Мадрид (Іспанія). Ці дані надаються Мадридською міською радою [10] на основі вільної ліцензії. Зокрема було використано дані про автомобільний трафік [11] та метеорологічні дані [12]. Ці дані представляють собою окремі файли, кожен з яких містить значення за кожен годину за всіма станціями.

Для проведення досліджень було використано дані за період з 1 січня 2019 року до 30 вересня 2022 року. За даний період діючими виявилися 59 станцій спостереження за трафіком, за якими і відбувалося створення моделей. Прогнозування реалізовувалося на 6 годин уперед, тобто на виході мало бути отримано 6 значень, кожне з яких відповідало кількості транспортних засобів, які подолали станцію спостереження протягом години. Оцінювання моделей виконувалося на основі середньої абсолютної

похибки або mean absolute error (MAE), середньоквадратичної похибки або mean square error (MSE), кореня середньоквадратичної похибки або root mean square error (RMSE) та коефіцієнта детермінації  $R^2$ .

Спочатку було виконано ряд експериментів, які стосувалися визначення тривалості вхідного часового інтервалу. Для цього було побудовано LSTM-моделі для кожної станції з довжиною вхідного часового ряду в 6 годин та 24 години. Відповідно було сформовано дві вибірки даних (перша з вхідним рядом з 6 значень і вихідним з 6 та друга з вхідним рядом з 24 значень і вихідним з 6 значень). Обидва варіанти LSTM-моделей мали одну вхідну ознаку з набором значень, кількість яких дорівнює описаній вище. Ця вхідна ознака визначалася даними трафіку за станцією, за якою відбувалося прогнозування автомобільного трафіку. У таблиці 1 представлено підсумкові результати, отримані з використанням побудованих описаним чином LSTM-моделей, що включають 2 приховані шари, відкидання між шарами та повнозв'язний шар. За тривалістю в 6 та 24 години було окремо побудовано 59 моделей для кожної станції. Далі ці значення були проаналізовані, а середні, мінімальні та максимальні отримані результати за кожним оціночним показником зведені до таблиці 1.

Таблиця 1

*Результати прогнозування автомобільного трафіку за станціями на основі LSTM-моделей з різним вхідним часовим інтервалом*

Модель прогнозування (тривалість вхідного ряду, годин)	Підсумкова оцінка	MSE	MAE	RMSE	$R^2$
LSTM (6)	Середня	0,002908	0,032192	0,049139	0,818415
	Мінімальна	0,000042	0,004379	0,006519	0,460282
	Максимальна	0,01016	0,066496	0,100799	0,926152
LSTM (24)	Середня	0,00215	0,028496	0,04205	0,854499
	Мінімальна	0,00004	0,004435	0,006304	0,403747
	Максимальна	0,007369	0,059623	0,085844	0,955042

З таблиці 1 видно, що використання вхідного часового ряду у 24 години замість 6 з подальшим прогнозуванням автомобільного трафіку на 6 наступних годин дозволило отримати кращі підсумкові результати. Так, в середньому значення MSE вдалося покращити на 26,07 %, MAE – на 11,48 %, RMSE – на 14,43 %, а коефіцієнта детермінації – на 4,41 %. Тобто за всіма показниками побудовані на основі часового ряду в 24 години результати виявилися кращими порівняно з використанням 6 годин на вхід. Окрім того можна помітити, що в найкращому випадку для однієї зі станцій вдалося отримати за цими 2 варіантами приблизно співставні результати, а у випадку найгіршого результату для однієї зі станцій MSE, RMSE та MAE, що дозволяють оцінити точність результатів, виявилися кращими для варіанту LSTM-моделі з 24 годинами при деякому зменшенні інформативності найгіршої з моделей.

Водночас додавання вхідних ознак на основі метеорологічних даних в середньому до покращення результатів не привезли, тобто дані виявилися надмірними, тому вони не були представлені в підсумку.

Далі було виконано дослідження прогнозування автомобільного трафіку за допомогою додавання додаткових ознак на основі даних трафіку інших станцій. LSTM-модель була побудована фактично в результаті з 3 вхідними ознаками: за станцією прогнозування та 2 станціями, відібраними зі всіх інших 4 способами. Перші 3 способи полягали у використанні коефіцієнта кореляції Пірсона на основі:

- значення наступної (першої майбутньої) години для станції прогнозування та всіх інших;
- значення наступної години для станції прогнозування та першої попередньої для всіх інших;
- значення всіх наступних годин для станції прогнозування та всіх минулих для всіх інших.

Четвертий спосіб полягав у використанні ансамблів дерев рішень, побудованих методом Random Forest. Використовувалися значення майбутньої години для станції прогнозування та попередньої для всіх інших. Підсумкові результати наведені в таблиці 2.

Таблиця 2

*Результати прогнозування автомобільного трафіку за станціями на основі LSTM-моделей з різними методами відбору вхідних ознак*

Модель прогнозування (метод відбору ознак)	Підсумкова оцінка	MSE	MAE	RMSE	$R^2$
1	2	3	4	5	6
LSTM (коефіцієнт кореляції Пірсона, поточна година)	Середня	0,002524	0,030302	0,045773	0,834732
	Мінімальна	0,000047	0,005028	0,006832	0,426328
	Максимальна	0,007229	0,05936	0,085022	0,948103
LSTM (коефіцієнт кореляції Пірсона, попередня година)	Середня	0,002439	0,030147	0,045108	0,839787

1	2	3	4	5	6
	Мінімальна	0,000043	0,004871	0,006552	0,517607
	Максимальна	0,007327	0,05901	0,085596	0,955723
LSTM (коефіцієнт кореляції Пірсона, всі наступні та попередні години)	Середня	0,002653	0,030788	0,046628	0,831158
	Мінімальна	0,000044	0,004801	0,006606	0,504903
	Максимальна	0,008727	0,064722	0,09342	0,955843
LSTM (Random Forest)	Середня	0,002356	0,028937	0,043898	0,851874
	Мінімальна	0,00004	0,00444	0,006336	0,522797
	Максимальна	0,009264	0,068737	0,096249	0,946106

За даними таблиці 2 помітно, що найкращі результати в середньому дозволили отримати LSTM-моделі з 3 вхідними ознаками, відібраними за допомогою методу Random Forest, на основі 6 годин спостереження. Такий спосіб у підсумку дозволив покращити значення точності результатів за показником MSE на 18,98 %, за показником MAE – на 10,11 %, за показником RMSE – на 10,67 %, а інформативності – на 4,09 % відносно LSTM, побудованої на основі 1 ознаки за 6 годин (таблиця 1). Всі інші способи призвели до отримання гірших результатів. Хоча при цьому обчислення коефіцієнтів кореляції вимагає значно меншого часу порівняно з Random Forest. Окрім того результати вказують також на те, що у випадку необхідності виконання швидких обчислень коефіцієнт кореляції Пірсона варто використовувати при визначенні кореляції між першою майбутньою годиною та останньою минулою, тобто з кроком у 1 годину. Необхідності обчислювати всі комбінації між входами та виходами немає, адже такий спосіб продемонстрував найменші покращення порівняно з LSTM з 1 входом та 6 годинами вхідних даних.

**Висновки та перспективи подальших досліджень.** Проблема прогнозування автомобільного трафіку є практично значущою і була розв’язана в даній роботі в розрізі прогнозування на 6 наступних годин. Було проведено експериментальне дослідження використання LSTM-моделей на основі даних, зібраних за 59 станціями. Отримані результати продемонстрували, що використання даних про трафік на станції прогнозування за минулі години для формування моделлю прогнозу на наступні години значно покращується за збільшення тривалості вхідного часового ряду з 6 годин до 24. Також до значного покращення призводить використання додаткових вхідних ознак на основі даних трафіку з інших станцій при виконанні відбору таких станцій. Найкращі результати продемонструвало використання ансамблів дерев рішень на основі методу Random Forest. Загальні результати дозволяють рекомендувати створені таким чином моделі для практичного використання зокрема при розробці прикладних програм.

У подальших дослідженнях необхідно проаналізувати створення моделей, які дозволяють інтегрувати отримані результати шляхом об’єднання додаткових вхідних ознак за даними трафіку з інших станцій та збільшення часового ряду за всіма вхідними ознаками одночасно, а також використовуючи більш складні структурні рішення LSTM-моделей.

#### References:

1. United Nations, Department of Economic and Social Affairs (2019), *World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)*, New York, United Nations, 126 p., [Online], available at: <https://population.un.org/wup/publications/Files/WUP2018-Report.pdf>.
2. Chen, R. & Yao, H. (2023), «Hybrid Graph Models for Traffic Prediction», *Applied Sciences*, No. 13 (15), doi: 10.3390/app13158673.
3. Hu, X., Liu, T., Hao, X. and Lin, C. (2022), «Attention-based Conv-LSTM and Bi-LSTM networks for large-scale traffic speed prediction», *The Journal of Supercomputing*, No. 78, pp. 12686–12709, doi: 10.1007/s11227-022-04386-7.
4. Tytarenko, V. & Shumliakivskyi, V. (2023), «Measurement and analysis of traffic delays in front of controlled intersections based on quadcopter images», *Technical Engineering*, No. 1 (91), pp. 228–234, doi: 10.26642/ten-2023-1(91)-228-234.
5. Fu, R., Zhang, Z. & Li, L. (2016), «Using LSTM and GRU neural network methods for traffic flow prediction», *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, Wuhan, China, CAA, pp. 324–328, doi: 10.1109/YAC.2016.7804912.
6. Abduljabbar, R.L., Dia, H., Tsai, T.-W. (2021), «Development and evaluation of bidirectional LSTM freeway traffic forecasting models using simulation data», *Scientific Reports*, No. 11, doi: 10.1038/s41598-021-03282-z.
7. Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y. & Liu, J. (2017), «LSTM network: a deep learning approach for short-term traffic forecast», *IET Intelligent Transport Systems*, Volume 11, No. 2, pp. 68–75.

8. Poonia, P. & Jain, V. (2020), «Short-Term Traffic Flow Prediction: Using LSTM», *2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)*, Lakshmanarath, Mody University of Science and Technology, pp. 1–4, doi: 10.1109/ICONC345789.2020.9117329.
9. Aggarwal, C. (2018), «Neural Networks and Deep Learning: A Textbook», *Springer*, 520 p., doi: 10.1007/978-3-319-94463-0.
10. *En portada – Portal de datos abiertos del Ayuntamiento de Madrid*, [Online], available at: <https://datos.madrid.es/portal/site/egob>.
11. *Aforos de tráfico en la ciudad de Madrid permanentes – Portal de datos abiertos del Ayuntamiento de Madrid*, [Online], available at: <https://datos.madrid.es/sites/v/index.jsp?vgnextoid=fabfb3e1de124610VgnVCM2000001f4a900aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD>.
12. *Datos meteorológicos. Datos horarios desde 2019 – Portal de datos abiertos del Ayuntamiento de Madrid*, [Online], available at : <https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9f8e4b2e4b284f1a5a0/?vgnextoid=fa8357cec5efa610VgnVCM1000001d4a900aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnnextfmt=default>.

**Льовкін** Валерій Миколайович – кандидат технічних наук, доцент, доцент кафедри програмних засобів Національного університету «Запорізька політехніка».

<https://orcid.org/0000-0002-6890-2807>.

Наукові інтереси:

- інформаційні технології;
- інтелектуальні системи;
- системний аналіз.

**Lovkin V.M.**

#### **Vehicle traffic forecasting models based on LSTM in application development**

The paper investigates the problem of traffic prediction. The issue is important for residents of modern cities and their groups while planning transportations in a city in particular and their own activities in general, taking into account that increased traffic leads to the increase in air pollution. The objective of the research was to study the influence of different input features of an LSTM model and the length of input time periods on the results of traffic prediction in a city. The proposed solutions make it possible to overcome the limitations concerning requirements for a whole system of observation stations and for a large set of historical data. These requirements are common for models proposed in the researches of other authors. In the paper the problem of traffic prediction, which is a matter of a time series forecasting, is considered for the horizon of the next 6 hours with the integration of the created models into applied programs, for which the appropriate data collection and processing procedures are defined. Experimental investigation was conducted using data collected at 59 observation stations and its results were evaluated based on a number of indicators (mean absolute, mean square and root mean square error) of accuracy and on the coefficient of determination for informativeness. The proposed prediction models are built based on long short-term memory. Experimental investigation confirmed the improvement in terms of accuracy and informativeness of the created models with a longer input time interval (24 hours instead of 6) and with additional input features based on traffic data of observation stations, which were selected from the whole set of stations using ensembles of decision trees based on the Random Forest method.

**Keywords:** traffic; prediction; time series; machine learning; long short-term memory.

Стаття надійшла до редакції 20.10.2023.