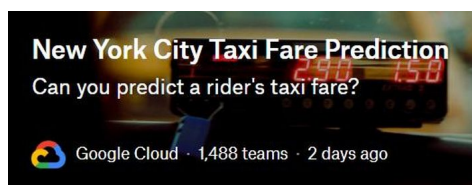
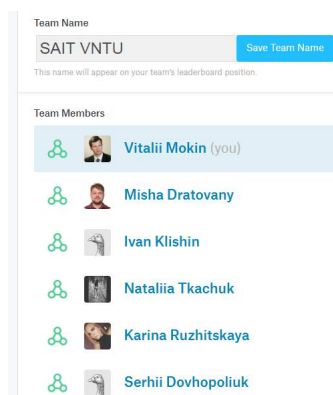


Завершилися чергові змагання з технологій машинного навчання на Kaggle, що є відомою у світі платформою змагань зі штучного інтелекту, системного аналізу та передбачувального моделювання (AI, Machine Learning, Data Science, Data Mining та ін.)

Призовий фонд деяких змагань сягає 1,5 млн. \$.



У цьогорічних змаганнях «New York City Taxi Fare Prediction» дебютувала команда «SAIT VNTU» системних аналітиків з ВНТУ – студентів різних курсів (Іван Клішин, [Каріна Ружицька](#), Наталія Ткачук) та аспірантів (Михайло Дратованій, Сергій Довгополюк) спеціальності 124 – «Системний аналіз» під керівництвом завідувача кафедри системного аналізу, комп'ютерного моніторингу та інженерної графіки, доктора технічних наук, професора Віталія Мокіна.



Методичну підтримку надавали фахівці відомої вінницької ІТ-фірми "Nestlogic", яка займається задачами у сфері штучного інтелекту (AI).

У конкурсі ( <https://www.kaggle.com/c/new-york-city-taxi-fare-prediction> ) був заданий тренувальний набір даних — дата, час і GPS-координати посадки в таксі, GPS-координати висадки з таксі, кількість пасажирів та вартість поїздок у жовтому таксі Нью-Йорка за 2009-2015 рр. (55 млн. рядків даних – csv-файл на 5,5 Гб).

Був заданий контрольний файл із 9914 поїздками (такі ж показники, але без вартості поїздки) і треба було передбачити якою буде вартість таких поїздок.

Ні тривалість поїздки, ні маршрут прямування не були відомими. І при цьому були випадки, коли координати збігались, тобто людина сіла, наприклад, на вокзалі, поїздила містом і на вокзалі вийшла та розплатилась за все одразу. І тепер відомо тільки дату і час посадки, однакові координати посадки і висадки та вартість поїздки – важки випадки для передбачення.

У тренувальному наборі даних були й відверто помилкові дані, які слід було знайти і відфільтрувати (переплутані широта і довгота, кількість пасажирів 0 або 30, від'ємна вартість поїздки та ін.).

Були рекомендації як вирішувати задачу, зокрема — використовувати методи XG-бустингу та мову програмування Python – це, власне, універсальна порада для усіх конкурсів Kaggle. А ще була порада спочатку пройти 36 он-лайн курсів спеціалізації Google Cloud, які є організаторами цього та й більшості курсів Kaggle.

У змаганні взяли участь майже півтори тисячі аналітиків та їх команд з усього світу.

«Ми приєднались до конкурсу за 3 тижні до його завершення і мали все швидко опанувати самі. Добре, що на кафедрі вже викладався Python та викладались інтелектуальні прикладні ІТ і можна було швидко розібратись що до чого. Розібрались. Зареєстрували команду. Спланували спільні дії. Ключове в задачах машинного навчання

– це пошук «features» (характеристик об'єктів та процесів, по яких будується модель). Одні займались інфраструктурою міста (аеропорти, метро, вокзали, порти, автобуси), інші – метеопараметрами, хтось – святковими днями. Хтось – вивчав форум та досвід інших. Вдалось придумати зовсім оригінальні рішення, які ніде не згадувались (кількість великих мостів чи тунелів на маршруті від місця посадки до місця висадки). Проведено системний аналіз об'єкту. Було зібрано та підключено з півсотні характеристик. Проведено дослідження їх впливу. Частина була забракована. По інших - ідентифікована модель XG-бустингу, яка дозволила досягти похибку у 3,2%. Для порівняння – останнє місце має похибку 39%, перше – 1,38%. 10-те місце – 2,82%, 100-те місце – 2,91%. Але потім, завдяки опануванню в останні дні ще й моделі LightGBM (високошвидкісний бустинг), вдалось поліпшити результат до 3,065% (320-те місце), який відрізняється на 1,68% від 1-го місця, на 0,25% від 10-го і на 0,16% від 100-го. Вдалось випередити 1168 команд. Враховуючи, що, через брак часу та обмеженість обчислювальних ресурсів, робота йшла тільки з 2 із 55 млн. рядків даних (3,6%), досягнута точність моделі є вражаючою», — розповідає професор Мокін.

Найближчим часом для студентів кафедри САКМІГ – спеціальності 124 і 126 буде проведена лекція по цих технологіях. Планується сформувати мінімум 3 команди під керівництвом викладачів кафедри, які викладають основні спецдисципліни на 124 та 126 ( <http://mmss.vntu.edu.ua/.../entrant-.../novi-distsiplini-kafedri>).

ІТ-компанія "Nestlogic" готова надавати призи студентам-переможцям конкурсів Kaggle з Вінниці.

Необхідні для вдалого виступу знання та навички будуть викладатись в межах дисциплін кафедри САКМІГ на різних курсах.

Все це дозволить підвищити якість навчання та конкурентоздатність випускників кафедри на ринку праці.

Ця діяльність є прикладом реалізації на кафедрі САКМІГ концепції проектного навчання ( <http://mmss.vntu.edu.ua/index.php/ua/project-edu> ).