

| Компетентності рівня<br>junior для Machine Learning  | Компанії                                      |        |             |
|--|---|--------|-------------|
|  | Загальна оцінка<br>компетентності<br>(max. 6) | CodeIT | GlobalLogic |
| <b>1. Основи машинного навчання</b>  |   |        |             |
| <b>1.1 Розуміння концепцій машинного навчання</b><br>Знає основні концепції машинного навчання, включаючи принципи роботи алгоритмів, типи даних та роль навчання у побудові моделей. Розуміє ключові етапи машинного навчання, такі як підготовка даних, навчання моделі, оцінка її якості та виведення результатів. Має базове розуміння процесу побудови моделей машинного навчання для різних типів завдань.   | 6   | 3      | 3           |
| <b>1.2 Види машинного навчання. Навчання з учителем (Supervised Learning). Навчання без вчителя (Unsupervised Learning). Навчання з підсиленням (Reinforcement Learning)</b><br>Розуміє процес навчання моделей на основі позначених даних і може використовувати алгоритми навчання з учителем (лінійна регресія, логістична регресія, дерева рішень, нейронні мережі) для класифікації та регресії. Має навички підбору параметрів моделей та оцінки їх точності. Вміє працювати з неозначеними даними, використовуючи алгоритми навчання без учителя (кластеризація, зниження вимірності) та знає, коли їх застосовувати. Розуміє принципи навчання з підсиленням і основні алгоритми (Q-learning, SARSA) для оптимізації дій агента. | 5   | 3      | 2           |
| <b>2. Програмування та бібліотеки</b>  |   |        |             |
| <b>2.1 Розуміння основ програмування на Python</b><br>Має навички роботи з основними елементами Python: змінними, типами даних (числа, рядки, списки, кортежі, словники), операціями (арифметичними, логічними, порівняннями), контрольними структурами (умовними операторами, циклами) і структурами даних (списки, кортежі, словники, множини). Вміє створювати функції, використовувати вбудовані функції та модулі. Розуміє основи ООП (класи, об'єкти, наслідування, інкапсуляція, поліморфізм) і може створювати класи для моделювання реальних об'єктів. Дотримується принципів написання чистого коду (PEP 8).   | 6   | 3      | 3           |
| <b>2.2 Використання бібліотек для машинного навчання: Scikit-learn, TensorFlow або PyTorch</b><br>Має досвід використання бібліотек машинного навчання, таких як Scikit-learn, TensorFlow або PyTorch, для створення, тренування та оцінки моделей. Вміє налаштовувати параметри моделей, застосовувати різні алгоритми та виконувати їхню оптимізацію. Розуміє основні принципи роботи кожної з бібліотек і знає, коли і яку з них використовувати в залежності від типу задачі.  | 5   | 3      | 2           |
| <b>3. Обробка та аналіз даних</b>  |   |        |             |
| <b>3.1 Очищення та підготовка даних</b><br>Здатний виконувати очистку даних, видаляти шум, помилки, а також заповнювати або видаляти пропуски перед проведенням аналізу та навчання моделей. Володіє методами роботи з текстовими, числовими та категоріальними даними для підвищення якості вхідних даних.  | 4   | 3      | 1           |

|   |   |   |   |
|---|---|---|---|
| <b>3.2 Аналіз відсутніх значень</b><br>Вміє використовувати різні методи для обробки відсутніх значень у наборах даних, такі як видалення пропусків, заповнення середнім, медіаною або іншими статистичними методами. Розуміє, коли і які методи слід застосовувати для мінімізації впливу відсутніх даних на результати аналізу.   | 4 | 3 | 1 |
| <b>3.3 Нормалізація та стандартизація</b><br>Розуміє важливість нормалізації та стандартизації даних для покращення роботи моделей машинного навчання. Вміє застосовувати техніки масштабування, такі як мінімакс-скейлінг або стандартизація (z-score), для того щоб уникнути проблем із різними шкалами даних.  | 4 | 3 | 1 |
| <b>3.4 Трансформація категоріальних даних</b><br>Знає методи перетворення категоріальних змінних у числові формати, такі як one-hot encoding або label encoding. Вміє застосовувати ці методи для коректної обробки категоріальних даних в алгоритмах машинного навчання.   | 4 | 3 | 1 |
| <b>3.5 Використання бібліотек pandas та numpy</b><br>Володіє навичками роботи з бібліотеками pandas та numpy для ефективною обробки, аналізу та маніпуляції даними. Вміє використовувати функції для обробки таблиць, масивів, маніпулювання стовпцями та рядками, а також виконувати математичні операції над даними.  | 5 | 3 | 2 |
| <b>3.6 Візуалізація даних допомогою бібліотек Matplotlib або Seaborn.</b><br>Вміє створювати інформативні графіки та діаграми для візуалізації даних за допомогою бібліотек Matplotlib або Seaborn. Використовує візуалізацію для виявлення закономірностей, трендів та аномалій у даних, а також для більш ефективною презентації результатів аналізу.   | 5 | 3 | 2 |
| <b>4. Моделі та методи машинного навчання</b>   |   |   |   |
| <b>4.1 Реалізація моделей класифікації та регресії</b><br>Здатний реалізувати моделі для класифікації та регресії, розуміє їх різницю: класифікація використовується для розподілу даних на категорії (логістична регресія, дерева рішень, SVM), а регресія — для прогнозування безперервних значень (лінійна, поліноміальна регресія). Вміє працювати з навчальними даними, налаштовувати моделі, проводити їх валідацію та оцінювати результати за метриками, такими як точність (accuracy), F1-score для класифікації та середня квадратична похибка (MSE) для регресії. | 5 | 3 | 2 |
| <b>4.2 Налаштування гіперпараметрів моделей. Grid search. Random search.</b><br>Розуміє важливість налаштування гіперпараметрів для підвищення точності моделей. Вміє застосовувати методи Grid search та Random search для пошуку оптимальних гіперпараметрів: Grid search проводить вичерпний пошук, а Random search випадково вибирає комбінації, що ефективніше для великих даних. Знає, як використовувати ці методи для оптимізації моделей і покращення їх продуктивності.   | 4 | 2 | 2 |
| <b>4.3 Кластеризація та сегментація</b><br>Розуміє методи кластеризації, які використовуються для групування подібних об'єктів у кластери на основі схожості даних. Вміє застосовувати алгоритми, такі як k-means, ієрархічна кластеризація та DBSCAN для сегментації наборів даних. Знає, як інтерпретувати результати кластеризації та використовувати їх для виявлення структур у великих масивах даних.   | 5 | 3 | 2 |
| <b>4.4 Зменшення розмірності</b><br>Здатний застосовувати техніки зменшення розмірності для спрощення моделі, зменшення часу обробки даних та полегшення візуалізації. Володіє навичками використання методів, таких як головні компоненти (PCA) або t-SNE, які дозволяють знизити кількість ознак у даних, зберігаючи їх важливі характеристики. Розуміє, як правильно застосовувати ці методи для зменшення шуму у високорозмірних даних та уникнення переобучення.   | 4 | 2 | 2 |
| <b>4.5 Виявлення аномалій</b><br>Вміє використовувати алгоритми для виявлення аномалій у наборах даних, що дозволяє знаходити відхилення або нестандартні патерни. Знає методи, такі як ізольовані ліси (Isolation Forests), метод локального фактору аномальності (LOF) та методи на основі статистичних тестів. Вміє застосовувати ці підходи для виявлення шахрайства, діагностики помилок у технічних системах або інших сфер, де важливе своєчасне визначення аномальних випадків.   | 4 | 2 | 2 |
| <b>5. Оцінка моделей</b>  |   |   |   |

|   |   |   |   |
|---|---|---|---|
| <p><b>5.1 Cross-Validation</b><br/>Знає методи крос-валідації для оцінки продуктивності моделей і запобігання переобученню. Вміє застосовувати k-fold крос-валідацію та leave-one-out валідацію для менших наборів даних. Розуміє, як інтерпретувати результати і приймати рішення про доопрацювання моделей на їх основі.</p>  | 4 | 2 | 2 |
| <p><b>5.2 Методи регуляризації. L1 та L2 регуляризація. Elastic Net.</b><br/>Розуміє регуляризацію як метод запобігання переобученню. Вміє застосовувати L1 регуляризацію (Lasso) для зменшення кількості параметрів, L2 (Ridge) для мінімізації їхніх значень, і Elastic Net, що поєднує обидва методи. Знає, як підбирати коефіцієнти регуляризації для балансу між підгонкою та узагальненням моделі.</p>  | 4 | 2 | 2 |
| <p><b>5.3 Метрики оцінки моделей: точність, F1-оцінка, AUC, RMSE та MAE.</b><br/>Знає, як використовувати різні метрики для оцінки якості моделей залежно від завдання. Сприймає, що для класифікаційних задач застосовується точність (accuracy), F1-оцінка (F1-score), AUC (Area Under the Curve). Для регресійних задач застосовується RMSE (Root Mean Squared Error), а також MAE (Mean Absolute Error), яка показує середню величину помилки в прогнозах моделі. Розуміє, як правильно обирати відповідні метрики залежно від типу моделі (класифікація чи регресія) та специфіки даних.</p>   | 5 | 2 | 3 |
| <b>6. Глибинне навчання та нейронні мережі</b>  |   |   |   |
| <p><b>6.1 Реалізація нейронних мереж. Перцептрони. Згорткові мережі (CNN). Рекурентні мережі (RNN).</b><br/>Знає основи побудови нейронних мереж, включаючи перцептрон як базовий елемент складніших архітектур. Розуміє, як використовувати згорткові нейронні мережі (CNN) для обробки зображень та інших даних зі структурою, автоматично виділяючи важливі ознаки. Також володіє навичками роботи з рекурентними нейронними мережами (RNN) для обробки послідовних даних, таких як текст чи часові ряди. Знає різновиди RNN, зокрема LSTM та GRU, які ефективно зберігають довготривалі залежності в даних.</p>   | 3 | 1 | 2 |
| <p><b>6.2 Трансферне навчання. Попередньо навчені моделі (VGG, ResNet, Inception).</b><br/>Знає концепцію трансферного навчання, що дозволяє використовувати попередньо навчені моделі для нових задач із меншими наборами даних. Вміє працювати з моделями, такими як VGG, ResNet та Inception, які є стандартом у глибинному навчанні. VGG забезпечує просту архітектуру для розпізнавання образів, ResNet вирішує проблему зникання градієнтів у глибоких мережах, а Inception оптимізує продуктивність через паралельні згорткові шари. Має навички модифікації цих моделей для конкретних завдань, зберігаючи частину навчених вагів і адаптуючи інші шари до нових даних.</p> | 3 | 1 | 2 |
| <b>7. Інтеграція моделей машинного навчання</b>   |   |   |   |
| <p><b>7.1 Методи NLP (Natural Language Processing)</b><br/>Розуміє основні концепції обробки природної мови, такі як токенізація, стемінг, лемматизація.</p>  | 5 | 3 | 2 |
| <p><b>7.2 Колаборативна фільтрація. Контентна фільтрація. Змішані системи</b><br/>Розуміє основи колаборативної фільтрації, включаючи матричну факторизацію та алгоритми на основі сусідів. Вміє реалізувати методи контентної фільтрації, що використовують атрибути об'єктів для рекомендацій. Знає, як поєднувати різні підходи (колаборативну та контентну фільтрацію) для створення змішаних рекомендаційних систем.</p>   | 3 | 1 | 2 |
| <p><b>7.3 Використання бібліотек Surprise та LightFM</b><br/>Володіє навичками роботи з бібліотеками Surprise та LightFM для створення рекомендаційних систем, налаштування і оцінки моделей за метриками RMSE та MAE. Вміє використовувати LightFM для побудови гібридних систем, що враховують як поведінкові дані користувачів, так і атрибути контенту. Здатний налаштувати гіперпараметри та покращувати точність рекомендацій, враховуючи рейтинги і контекстуальні ознаки.</p>   | 2 | 1 | 1 |
| <p><b>7.4 Взаємодія з вебдодатками (Flask, Django)</b><br/>Знає, як інтегрувати моделі машинного навчання з вебдодатками, використовуючи Flask та Django. Розуміє принципи роботи обох фреймворків. Вміє налаштувати RESTful API для надання сервісів, що дозволяє користувачам взаємодіяти з моделями машинного навчання через HTTP-запити. Знає, як реалізувати ендпоінти для отримання даних, запуску прогнозів та повернення результатів. Має навички з обробки запитів і відповідей у форматах JSON.</p>   | 5 | 3 | 2 |

|   |   |   |   |
|---|---|---|---|
| <b>7.5 Розгортання моделей на вебсерверах</b><br>Вміє розгортати моделі машинного навчання на веб-серверах, використовуючи різні платформи хостингу. Знає, як забезпечити доступність моделей для користувачів через інтерфейс або API, включаючи налаштування серверів та управління ресурсами.  | 3 | 2 | 1 |
| <b>7.6 Використання Docker для управління моделями</b><br>Розуміє, використання Docker для контейнеризації моделей машинного навчання, спрощуючи їх розгортання та управління залежностями. Вміє створювати образи, налаштовувати контейнери, використовувати Docker Compose для багатоконтейнерних додатків. Знає, як упаковувати моделі та бібліотеки в Docker-образи для консистентності середовищ і масштабування розгортань, включаючи хмарні платформи. | 4 | 2 | 2 |
| <b>7.7 Автоматизація процесу розробки моделей (AutoML)</b><br>Знає, як використовувати AutoML для автоматизації вибору алгоритмів, налаштування гіперпараметрів і оцінки моделей. Володіє інструментами, такими як TPOT, H2O.ai, AutoKeras, та інтегрує AutoML у робочий процес, приймаючи обґрунтовані рішення на основі отриманих результатів.  | 2 | 1 | 1 |
| <b>7.8 Інструменти для роботи з великими обсягами даних</b><br>Вміє працювати з екосистемою Hadoop, включаючи HDFS (Hadoop Distributed File System) для зберігання великих обсягів даних і MapReduce для їх обробки. Знає Apache Spark для обробки даних в пам'яті, що дозволяє швидше аналізувати великі набори даних у порівнянні з традиційними методами, такими як Hadoop MapReduce.  | 2 | 1 | 1 |
| <b>7.9 Архітектура потокової обробки даних</b><br>Розуміє, як використовувати Kafka для розподіленої обробки великих обсягів даних з низькою затримкою. Знає Flink для складної обробки потоків у реальному часі, Storm для обробки даних з високою пропускнуною здатністю та низькою затримкою, і Spark Streaming для аналізу даних у реальному часі з інтеграцією з іншими компонентами Spark.  | 2 | 1 | 1 |
| <b>8. Розмовна англійська</b>   |   |   |   |
| <b>8.1. Правильна побудова фрази з узгодженням часу</b><br>Вміє використовувати відповідні часові форми дієслів і структури речень для точного вираження часу в комунікації. Здатний правильно використовувати часові форми, такі як Present Simple, Present Continuous, Past Simple, Past Continuous, Future Simple, для передачі правильного значення в повідомленні.   | 4 | 2 | 2 |
| <b>8.2 Професійна IT термінологія в розмові з замовником</b><br>Знає і використовує професійні IT-терміни та скорочення, які використовуються в IT-індустрії, для чіткого та зрозумілого спілкування зі замовником. Вміє перекладати складні технічні концепції на доступну мову для замовника, не знайомого з IT-галуззю.  | 4 | 2 | 2 |
| <b>8.3 Професійна IT термінологія у діловому звіті</b><br>Знає і використовує спеціалізовану IT-термінологію та фразеологію при написанні ділових звітів. Вміє передати інформацію про технічні питання та проекти зрозуміло та конкретно, використовуючи адекватні IT-терміни та аббревіатури.   | 5 | 3 | 2 |
| <b>8.4 Професійна термінологія для Machine Learning</b><br>Знає та розуміє специфічну термінологію, що використовується в контексті Machine Learning. Розуміє та використовує специфічні терміни, патерни, алгоритми, фреймворки, бібліотеки та інші інструменти, які використовуються в Machine Learning.  | 4 | 2 | 2 |
| <b>8.5 Неформальне спілкування (small talks)</b><br>Вміє вести неформальну розмову, розпочинати діалоги та підтримувати невимушену атмосферу. Знає загальноживану термінологію та фрази, які використовуються у неформальних розмовах, таких як загальні питання про погоду, цікаві події, особисті захоплення, спорт, фільми тощо.   | 5 | 3 | 2 |

## Коментарі до таблиці

Оцінки компетентностей вказані станом на 2024/2025 рр., по шкалі від 0 до 3.

- «3» - компетентність вкрай важлива, якщо кандидат нею не володіє, його подальше просування на вакансію
- «2» - компетентність, знання якої обов'язково знадобиться кандидату при подальшому професійному зростанні.
- «1» - некритична компетентність, якою можна знехтувати.
- «0» - неактуальна компетентність.