Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет
имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий - РТФ

Базовая кафедра «Аналитика больших данных и методы видеоанализа»

ДОПУСТИТЬ К ЗАЩИТЕ ПЕРЕД ГЭК

/ Зав. кафедрой «АБД и МВ»

_____ М.А. Медведева
(подпись) (Ф.И.О.)
02.06.2023

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

«Разработка инструмента для контроля производственных процессов»

Научный руководитель: Медведев М.А.
доцент, к. э.н.
Научный руководитель: Медведева М.А.
Доцент, к.ф.-м.н.

Студент группы РИМ-210980 Ярославский К.А.

Екатеринбург
2023

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

РЕФЕРАТ

Тема магистерской диссертации:

«Разработка инструмента для контроля производственных процессов»

Магистерская диссертация выполнена на 89 страницах, содержит 2 таблицы, 17 рисунков, 60 использованных источников.

Каждое производство, как малые развивающиеся фирмы, так и большие, уже занявшие позиции на рынке, нацелены на рост прибыли и закрепление своих позиций. Зачастую производственный процесс состоит из множества звеньев, и приостановка даже одного из них будет тормозить всю цепочку и в итоге может негативно сказаться на итоговом продукте и в потенциале привести к серьёзным убыткам. При этом аномалии, приводящие к остановке рабочего процесса, не всегда происходят единовременно – многие неполадки оборудования начинают малозаметно сказываться на итоговом продукте задолго до выхода оборудования из строя. Также на любой производственный процесс могут воздействовать факторы, которые не препятствуют достижению поставленных целей и выполнению задач, но понижают эффективность производственного процесса. Для поддержания непрерывности производственного процесса необходимо иметь способы поиска аномалий в процессе для своевременного их устранения.

В работе поставлена цель: создание программного продукта, который можно применить для контроля производственных процессов с использованием статистических методов.

Для достижения данной цели необходимо выполнить следующие задачи:

1. Изучить статистические методы контроля производственных процессов.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

2. Изучить существующие реализации методов контроля производственных процессов.

3. Сформулировать требования к разрабатываемому инструменту

4. Сформировать стек технологий.

5. Создать интерактивную форму выбора данных из предоставленной заказчиком базы данных.

6. Разработать функции ПО для визуализации статистических показателей процесса.

Объект исследования: статистические методы контроля производственных процессов.

Предмет исследования: разработка ПО для статистического анализа производственных данных.

Новизна работы заключается в использовании статистических методов контроля качества и открытой аналитической системы Jupyter Lab для контроля стабильности рабочих процессов на предприятии и своевременного обнаружения аномалий.

Практическая значимость работы – это внедрение разработанного инструмента в более масштабную информационную систему EXPERT BASE, которая будет применяться на Абинском Электрометаллургическом Заводе для непрерывного контроля производственных процессов.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

СОДЕРЖАНИЕ

ВВЕДЕНИЕ....6
1 ОБЗОР И АНАЛИЗ МЕТОДОВ КОНТРОЛЯ ПРОИЗВОДСТВЕННЫХ ПРОЦЕССОВ....9
1.1 Статистические методы контроля производственных процессов....9
1.2 Контрольные карты как метод контроля производственных процессов....22
1.3 Виды контрольных карт....25
1.4 Построение контрольных карт....26
1.5 Патентный анализ....31
1.6 Результаты и выводы по первой главе....33
2 ВЫБОР СТЕКА ТЕХНОЛОГИЙ ДЛЯ РАЗРАБОТКИ....34
2.1 Язык программирования....34
2.2 Среда программирования....35
2.3 Менеджер пакетов....36
2.4 Решение для соединения с базой данных....38
2.5 Решение для взаимодействия с базой данных....39
2.6 Инструмент для работы со статистическими функциями....50
2.7 Решение для работы с массивами данных....58
2.8 Решение для визуализации графиков....62
2.9 Решение для развёртывания формы....65
2.10 Система управления базами данных....66
2.11 Графический клиент для СУБД....68
2.12 Результаты и выводы по второй главе....68
3 РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ....70
3.1 Характеристика предприятия....70
3.2 Анализ требований заказчика....72
3.3 Планирование....73
3.4 Дизайн....76
3.5 Разработка....78
3.6 Расчёт экономической эффективности....81
3.7 Результаты и выводы по третьей главе....83
ЗАКЛЮЧЕНИЕ....84

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....86

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

ВВЕДЕНИЕ

Каждое производство, как малые развивающиеся фирмы, так и большие, уже занявшие позиции на рынке, нацелены на рост прибыли и закрепление своих позиций. Зачастую производственный процесс состоит из множества звеньев, и приостановка даже одного из них будет тормозить всю цепочку и в итоге может негативно сказаться на итоговом продукте и в потенциале привести к серьёзным убыткам. При этом аномалии, приводящие к остановке рабочего процесса, не всегда происходят единовременно – многие неполадки оборудования начинают малозаметно сказываться на итоговом продукте задолго до выхода оборудования из строя. Также на любой производственный процесс могут воздействовать факторы, которые не препятствуют достижению поставленных целей и выполнению задач, но понижают эффективность производственного процесса. Для поддержания непрерывности производственного процесса необходимо иметь способы поиска аномалий в процессе для своевременного их устранения.

В работе поставлена цель: создание программного продукта, который можно применить для контроля производственных процессов с использованием статистических методов.

Для достижения данной цели необходимо выполнить следующие задачи:

1. Изучить статистические методы контроля производственных процессов.
2. Изучить существующие реализации методов контроля производственных процессов.
3. Сформулировать требования к разрабатываемому инструменту
4. Сформировать стек технологий.
5. Создать интерактивную форму выбора данных из предоставленной заказчиком базы данных.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

6. Разработать функции ПО для визуализации статистических показателей процесса.

Объект исследования: статистические методы контроля производственных процессов.

Предмет исследования: разработка ПО для статистического анализа производственных данных.

В работе используются следующие методы исследования:

- теоретические (анализ и синтез, сравнение, дедукция, моделирование);
- эмпирические (эксперимент, измерение, математический анализ).

Анализ библиометрических показателей за последние четыре года (2019-2022) показывает стабильный рост количества публикаций в конкретно очерченной области исследования по всему миру. Так, в 2019 по всему миру было опубликовано 132 публикации в данной области, а в 2022 – 220 публикаций (цифра основывается на неполных данных о числе опубликованных статей, в реальности это число может оказаться ещё больше). Отдельно стоит выделить стабильность интереса к данной тематике в регионах Европы и России.

Новизна работы заключается в использовании статистических методов контроля качества и открытой аналитической системы Jupyter Lab для контроля стабильности рабочих процессов на предприятии и своевременного обнаружения аномалий.

Практическая значимость работы – это внедрение разработанного инструмента в существующую информационную систему EXPERT BASE, которая будет применяться на производствах для непрерывного контроля производственных процессов.

Эмпирическая база исследования: федеральный государственный стандарт высшего образования, ГОСТы, электронные источники, официальный сайт компании «ДАТА-ЦЕНТР Автоматика».

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Работа состоит из трёх глав: Обзор и анализ статистических методов контроля производственных процессов; Методология исследования; Разработка программного обеспечения. Первая глава состоит из четырёх подпунктов: Статистические методы контроля производственных процессов; Контрольные карты как метод контроля производственных процессов; Виды контрольных карт; Построение контрольных карт; Патентный анализ.

Вторая глава включает в себя: Выбор языка программирования; Выбор среды программирования; Выбор менеджера пакетов; Выбор решения для соединения с базой данных; Выбор решения для взаимодействия с базой данных; Выбор инструмента для работы со статистическими функциями; Выбор решения для работы с массивами данных; Выбор решения для визуализации графиков; Выбор решения для развёртывания формы; Система управления базами данных; Графический клиент для СУБД.

Третья глава имеет следующую структуру: Постановка задачи; Анализ требований заказчика; Планирование; Дизайн; Разработка; Расчёт экономической эффективности.

Каждая глава имеет отдельный подпункт с выводами о главе.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

1 ОБЗОР И АНАЛИЗ МЕТОДОВ КОНТРОЛЯ ПРОИЗВОДСТВЕННЫХ ПРОЦЕССОВ

1.1 Статистические методы контроля производственных процессов

Современные производственные процессы требуют обеспечения высокой однородности качества конечной продукции, а также того, чтобы уровень конечной продукции, несоответствующей стандартам качества, был максимально низким [1]. Существует множество методов для того, чтобы оценить ожидаемое качество конечной продукции. Одним из таких методов является оценка показателей возможностей процессов. Метод оценки показателей возможностей процессов может быть применён для широкого спектра уровней несоответствий конечной продукции стандартам качества.

Ещё при проектировании продукции и процессов рекомендуется проводить оценку ожидаемого качества. Эта оценка должна производиться на основе потенциальных характеристик процесса. Такой подход позволит избежать конфликта между требованиями, указанными в технологических или же конструкторских документах, и возможностями реальных процессов во многих случаях.

Принятие экономически верных решений относительно связанных с процессом действий – это основная цель системы по управлению любым производственным процессом. Принятие таких решений требует соблюдения баланса между последствиями излишнего управления и недостаточного управления. Излишнее управление — это принятие не вполне необходимых решений, в то время как недостаточной управление – это непринятие необходимых решений [2]. Необходимо рассматривать данные риски в контексте изменчивости технологических процессов. Выделяют два типа таких причин: особые причин и обычные причин. При таком подходе основной критерий

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

процесса – это его устойчивость. При этом также важно учитывать, что устойчивость процесса напрямую зависит от причин его изменчивости.

Изменчивостью обычно называют различия, обнаруженные среди индивидуальных значений процесса [3]. Подобные различия неизбежно будут обнаружены в производственном процессе. Некоторые из причин изменчивости производственного процесса вызывают кратковременные различия между единицами продукции. В то же время некоторые причины способны создавать изменения в продукте на длительных временных промежутках.

Выделяют следующие причины изменчивости:

1. Специальная (особая) — внешние воздействия на систему, а также факторы, которые воздействуют на процесс нерегулярно. В случае, если все особые причины изменчивости процесса не были вовремя идентифицированы и устранены, их влияние скажется на выходе производственного процесса непредсказуемым образом. При наличии особых причин выход процесса не стабилен во времени.
2. Общая (обычная/случайная) — множественные источники изменчивости процесса, которые имеют стабильное и повторяющееся распределение во времени. Процесс, изменчивость которого определяется только общими причинами, находится в статистически устойчивом состоянии. Обычные причины ведут себя как устойчивая система случайных причин. Если присутствуют и не меняются только обычные причины, исход процесса предсказуем.

Процесс считается статистически стабильным, если источниками изменчивости являются только обычные причины [4]. Одной из функций системы управления процессом является обеспечение статистических сигналов при наличии особых причин вариации и недопущение ложных сигналов при отсутствии таких причин. Это позволяет предпринять соответствующие действия

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

в ответ на эти особые причины (устранить их или сохранить неизменными, если это выгодно).

Существует ряд показателей, которые используются для оценки возможностей процесса. Во-первых, процесс должен быть статистически стабильным, и, кроме того, значения отдельных процессов должны быть близки к нормальному распределению [5].

Показатели возможностей представляют собой потенциальную и фактическую способность процесса соответствовать заданным техническим критериям для количественно измеряемых значений выходных показателей качества.

Показатели возможностей используют для следующих целей:

- предконтрактный анализ потенциальных возможностей поставщика удовлетворять требования потребителя;
- установление в контрактах (договорах на поставку) требований к процессам;
- планирование качества разрабатываемой продукции;
- приемка процессов на основе опытных партий;
- аттестация процессов;
- планирование приемочного контроля;
- планирование непрерывного улучшения процессов;
- аудиты второй стороной и внутренние аудиты процессов.

Надежность используемого показателя возможностей процесса зависит от изменчивости результатов измерения, обусловленной используемой процедурой измерения. Здесь общая погрешность измерения ΣΔ должна составлять (Y3 .... Y5) T (допуска) контрольного параметра.

Присущая процессу изменчивость является частью изменчивости процесса, которая возникает только по нормальным причинам [6]. Эта изменчивость

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

оценивается по контрольному списку с помощью соотношения R/d2 или S/c4 (d2 и c4 — стандартные коэффициенты, зависящие от объема выборки).

Общая изменчивость процесса — это изменчивость, обусловленная нормальными и особыми причинами. Эта изменчивость оценивается с помощью стандартного отклонения выборок для всех индивидуальных значений.

Воспроизводимость процесса — это интервал внутренней изменчивости процесса kσ только для статистически стабильных процессов, где σ обычно оценивается как R/d2 (σR/d2), а k зависит от характера распределения и вероятности надежности показателя качества процесса [7].

Индекс воспроизводимости C_p определяется как отношение допуска к оцененной внутренней изменчивости процесса без учета выравнивания процесса (1)

$$Cp = \frac{(USL-LSL)}{k\sigma}. \qquad (1)$$

В данной формуле USL — верхняя граница поля допуска; LSL — нижняя граница поля допуска, σ – среднеквадратическое значение для R/d2, где R — среднее арифметическое нескольких выборок, d2 — коэффициент, зависящий от объёма выборки и закона распределения показателя качества процесса, k — коэффициент, зависящий от закона распределения показателя качества процесса и достоверной вероятности.

Обычно при расчете статистических показателей возможностей процесса принимают допущение, что показатель качества процесса распределен по нормальному закону распределения, и для вероятности P = 99,37 % коэффициент k принимают равным 6 [8].

Пригодность процесса – это интервал в kσ полной изменчивости, где σ обычно оценивается при помощи выборочного стандартного отклонения.

Индекс пригодности P определяется как допуск, делённый на оценку полной изменчивости процесса без учёта его центровки (2).

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

$$P = \frac{USL-LSL}{k\sigma}. \qquad (2)$$

Верхний индекс воспроизводимости CPU определяется как отклонение среднего уровня процесса от верхнего предела поля допуска, делённое на действительный верхний разброс процесса (3).

$$CPU = \frac{USL-X'}{0{,}5k\sigma}. \qquad (3)$$

Нижний индекс воспроизводимости CPL определяется как отклонение среднего уровня процесса от нижнего предела поля допуска, делённое на действительный нижний разброс процесса (4).

$$CPL = \frac{X'-LSL}{0{,}5k\sigma}. \qquad (4)$$

Индекс воспроизводимости C_{pk} учитывает центровку процесса и определяется как минимальное из CPU и CPL. Он связывает разность между средним процесса и ближайшим пределом поля допуска с половиной присущей процессу изменчивости [9].

По известным значениям C_p и C_{pk} можно определить интервал, в котором находится ожидаемый уровень несоответствий. По значению C_{pk} определяют максимально возможное значение ожидаемого уровня несоответствий, по значению C_p – минимально возможное. Вследствие использования выборочных оценок для получения C_p и C_{pk} и ограниченности объемов наблюдения полученные значения ожидаемых уровней несоответствий могут заметно отличаться от фактически наблюдаемых уровней несоответствий действующих процессов, поэтому значение ожидаемых уровней несоответствий используют только для предварительных оценок качества процессов и мониторинга улучшений [10].

Индекс пригодности P_{pk} учитывает центровку процесса и определяется как минимальное из PPU и PPL (Данные величины рассчитываются аналогично CPU и CPL, но вместо отклонения всего процесса берётся стандартное отклонение по выборке). Он связывает разность между средним процесса и ближайшим

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

пределом поля допуска с половиной полной изменчивости процесса. Данный показатель, как и индекс пригодности P_p, должен использоваться только для сравнения или вместе с C_p и C_{pk}, а также для измерения и выбора приоритетов усовершенствования во времени.

Индексы воспроизводимости C_{pk} и пригодности P_{pk} применяются при измерении результатов непрерывного усовершенствования с использованием временных трендов и при выборе приоритетного направления, в котором процессы должны совершенствоваться.

Для характеристики процесса также используют отношение воспроизводимости и отношение пригодности [11].

Ни один приведённый отдельный индекс не могут дать исчерпывающие описание процесса. Все возможные индексы рекомендуется рассматривать совместно. Все характеристики воспроизводимости должны относится к характеристике одного процесса. Никогда не следует объединять или усреднять результаты по воспроизводимости для нескольких процессов в один индекс [12].

Управляемым считается статистически стабильный процесс, индекс пригодности которого превышает единицу. В этом случае поле рассеяния показателя качества, отражающее изменчивость процесса меньше допуска.

Основной целью метода статистического приемочного контроля является обеспечение оценки качества продукции, отправленной на проверку, и четкого принятия результатов оценки качества продукции поставщиком и заказчиком.

В системе менеджмента качества производитель должен нести ответственность за качество продукции. Во время инспекции ответственность за качество может переходить от производителя к инспектору [13]. Это происходит, когда руководство просят разобраться с несоответствующей продукцией, а все, что происходит на производстве, выявляется в ходе инспекции. Такой подход может привести к ненужной работе, расходам и снижению качества как для

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

заказчика, так и для производителя. Если производитель не делает этого, то у контролера нет возможности произвести высококачественный продукт.

Преимущество этого типа контроля заключается в том, что ответственность за качество продукции возлагается на производителя, а контролер не играет никакой роли в выборе продукции [14]. Производитель должен доказать, что продукт качественный, иначе для того, чтобы забраковать партию, потребуются значительные усилия и затраты. Выборочный контроль снижает усилия и затраты на контроль и обеспечивает хорошее качество продукции для покупателя. Так как при статистическом приемочном контроле решение принять или отклонить партию продукции реализуется по результатам контроля выборки, то всегда имеется некоторая вероятность выбрать ошибочное решение. При этом имеется риск как поставщика, так и потребителя [15].

Под риском поставщика понимается вероятность отклонения партии продукции, обладающей приемлемым уровнем качества (нормативный документ ГОСТ Р 50779.11-2001) [16].

Риск потребителя — это вероятность приемки партии продукции, обладающей предельным уровнем качества (ГОСТ Р 50779.11-2001 трактует это понятие как при данном плане выборочного контроля вероятность приемки партии или процесса, когда их уровень качества имеет значение, признаваемое по плану неудовлетворительным, например, значение предельного уровня качества).

Управление приемкой основывается на результатах контроля единицы продукции. Под единицей контроля понимается: продукт, количество материала, услуга, деятельность или процесс, организация или человек, или их комбинация.

Статистический приемочный контроль может проводиться по количественным, качественным или альтернативным критериям [17].

Статистический приемочный контроль по количественным признакам — это проверка качества продукции, при которой определяются значения

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

контрольных параметров и принимаются последующие решения о контролируемом предприятии или процессе путем сравнения с критериями контроля (нормативный документ СТБ ГОСТ Р 50779.11-2001).

Статистический приемочный контроль с помощью признаков качества — это проверка качества продукции, при которой каждый элемент контроля относится к определенной группе и последующее решение о контролируемой совокупности или процессе принимается на основе соотношения количества элементов, обнаруженных в различных группах.

Альтернативой статистическому приемочному контролю является контроль качества продукции с помощью атрибутов качества, при котором каждая проверенная единица продукции классифицируется как соответствующая или несоответствующая, а последующее решение принимается в отношении контролируемой совокупности или процесса на основе количества несоответствующих единиц продукции, обнаруженных в выборке, или количества несоответствующих изделий на фиксированное количество продукции. Решение основывается на результатах сравнения с контрольным эталоном (нормативный документ ГОСТ Р 50779). Для изучения распределений случайных величин в математической статистике пользуются рядом числовых характеристик, определяющих положение центра группирования случайной величины и ее рассеивание около этого центра [18].

Числовые характеристики положения центра группирования носят общее название мер положения, а числовые характеристики рассеивания — мер рассеивания.

Для статистических оценок мер положения при теоретическом распределении является математическое ожидание [19]. При эмпирическом распределении же используются следующие меры:

- среднее арифметическое значение;
- среднее арифметическое взвешенное;

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

- среднее гармоническое;
- среднее геометрическое;
- среднее геометрическое взвешенное;
- среднее квадратическое;
- среднее квадратическое взвешенное;
- середина размаха;
- медиана;
- мода.

Математическим ожиданием E(X) дискретной случайной величины *X* называется сумма произведений возможных ее значений на соответствующие вероятности.

Математическое ожидание E(X) непрерывной случайной величины *X*, имеющей плотность вероятности *f(X),* рассчитывается как интеграл от произведений возможных ее значений на соответствующие вероятности, если интеграл сходится абсолютно.

Математическое ожидание непрерывной случайной величины определяется значительно сложнее с использованием интегрального исчисления.

Среднее арифметическое значение, среднее гармоническое, среднее геометрическое, и среднее квадратическое можно рассчитать по формуле среднего степенного.

Среднее арифметическое взвешенное, среднее геометрическое взвешенное, среднее квадратическое взвешенное можно рассчитать по формуле среднего взвешенного.

Средним арифметическим значением случайной величины называется отношение суммы всех значений случайной величины, полученных в результате конечного числа испытаний, к числу испытаний (*согласно СТБ ГОСТ Р 50779.10 среднее арифметическое – сумма значений, деленная на их число)*.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Среднее арифметическое получают путем подстановки в формулу среднего степенного показателя степени z, равного 1.

Вышеприведенные выражения справедливы при контроле показателей по количественному признаку.

При контроле показателя по альтернативному признаку оцениваемый показатель может принимать только два взаимоисключающих значения, которым сопоставляются два количественных значения: 1 и 0.Частостью варианта 1 (как правило, обозначается p) является доля единиц, обладающих данным признаком в общей статистической совокупности. Разность 1 – p = q является частностью варианта 0. Таким образом, среднее арифметическое при контроле по альтернативному признаку вычисляется как p.

Средним арифметическим взвешенным значением случайной величины называется сумма произведений значений случайной величины на их частности (*согласно СТБ ГОСТ Р 50779.10 взвешенное среднее арифметическое – сумма произведений каждого значения на его вес, деленная на сумму весов, где веса – неотрицательные коэффициенты, связанные с каждым значением).*

Для непрерывных случайных величин X_i — это середина равного интервала, который делит ряд значений X.

Взвешенное среднее арифметическое получается путем подстановки в формулу взвешенного среднего порядка z, равного 1.

Как правило, среднее арифметическое — это взвешенное среднее арифметическое.

Среднее гармоническое вычисляется путем подстановки в формулу среднего степенного порядка z, равного -1.

Среднее геометрическое — это корень n-й мощности из произведения значений случайных величин.

Среднее геометрическое получается путем подстановки в формулу среднего степенного порядка z, равного 0.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Взвешенное среднее геометрическое получается путем подстановки в формулу взвешенного среднего порядка z с порядком z, равным 0.

Среднее геометрическое взвешенное используется для анализа динамики явления и определения среднего темпа роста. При расчете средней геометрической индивидуальные значения случайных величин являются относительными показателями динамики, полученными как отношение каждого уровня ряда к предыдущему уровню.

Среднее квадратическое — это корень n-й мощности из произведения значений случайных величин. Корень среднего получается путем подстановки в формулу среднего степенного порядка z, равного 2.

Согласно закону Боярского о среднем значении единичной статистической совокупности, гармоническое среднее меньше геометрического среднего, которое в свою очередь меньше арифметического среднего, которое в свою очередь меньше квадратичного среднего.

Таким образом, значение среднего возрастает с увеличением экспоненты z.

Серединой размаха называют полусумму наибольшего и наименьшего значений *(согласно СТБ ГОСТ Р 50779.10 середина размаха – это среднее арифметическое между наибольшим и наименьшим наблюденными значениями количественного признака).*

Когда n значений меры расположены в порядке возрастания, наиболее центральное значение называется медианой (согласно СТБ ГОСТ Р 50779.10, медиана — это квадрат порядка p = 0,5). n — нечетное число, а медиана — значение, расположенное на 1/2(n + 1).

Мода — наиболее часто встречающееся значение в статистической совокупности (согласно СТБ ГОСТ Р 50779.10, мода — это значение случайной величины, для которого функция распределения массы вероятности или функция плотности вероятности максимальна).

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

В случае эмпирических распределений дискретных случайных величин максимальное значение моды получается непосредственно по классическому определению. Для эмпирического распределения непрерывных случайных величин сначала определяется интервал моды $h_k = x_k — x_{k-1}$, соответствующий максимальной частоте f_k. Значение моды в интервале моды определяется по интерполяционной формуле Р. М. Орженцкого.

Математическое ожидание обычно используется в качестве меры положения для теоретических распределений, в которых возможные значения *X* оцениваются при помощи вероятностей. В эмпирических распределениях, где наблюдаемые значения *X* оцениваются при помощи частот или частостей, в качестве меры положения используется среднее арифметическое, среднее арифметическое взвешенное, среднее гармоническое, среднее геометрическое, среднее геометрическое взвешенное, среднее квадратическое, среднее квадратическое взвешенное, середина размаха, медиана и мода.

Математическая статистика использует несколько числовых свойств для изучения распределения случайной величины, определяя местоположение центра скопления случайной величины и дисперсию вокруг этого центра.

Числовые свойства расположения центра скопления известны как мера расположения, а численные свойства дисперсии известны как мера дисперсии.

Статистические оценки меры дисперсии используются для дисперсии, коэффициента вариации и количественной оценки теоретического распределения, а также для стандартного отклонения и разброса эмпирического распределения.

Дисперсия дискретной случайной величины — это сумма квадратов произведений отклонений случайной величины X от ее математического ожидания, выраженных в терминах соответствующих вероятностей.

Дисперсия непрерывной случайной величины с плотностью вероятности p(X) вычисляется как произведение квадрата отклонения случайной величины X

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

от ее математического ожидания и соответствующей вероятности, если интеграл сходится.

Эта величина применяется в качестве меры рассеивания теоретического распределения, а для эмпирического распределения используется аналогичная величина σ^2, которая определяется как сумма произведений квадратов отклонений значений случайной величины X_i от ее среднего арифметического значения X на соответствующее частности $f_i/(n-1)$.

Приведенная выше формула справедлива при контроле показателей с количественным атрибутом.

При контроле показателей с альтернативными атрибутами оцениваемый показатель может принимать только два взаимоисключающих значения, сравниваемых с двумя количественными значениями, т.е. 1 и 0. Частота варианта 1 (обычно обозначается как p) — это доля единиц с данным признаком в общей статистической совокупности. Разность 1 - p = q — это частота варианта 0. Таким образом, дисперсия эмпирического распределения в альтернативном контроле признака рассчитывается как p*q.

Таким образом, дисперсия эмпирического распределения случайной величины в альтернативном контроле признака равна произведению доли единиц с этим признаком и доли единиц без этого признака.

Когда p и q равны, т.е. p = q = 0,5, дисперсия эмпирического распределения случайной величины, контролирующей альтернативный признак, имеет максимальное значение pq = 0,25.

На практике используется не сама дисперсия, а ее квадратный корень, который известен как стандартное отклонение (стандартное отклонение). Размерность σ совпадает с размерностью самой случайной величины X.

Коэффициентом вариации называют отношение стандартного отклонения случайной величины к ее математическому ожиданию.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Квантилем z случайной величины X называется такое значение случайной величины, которому соответствует значение интегральной функции распределения, равное z. (согласно СТБ ГОСТ Р 50779.10 квантиль – это значение случайной величины Xp, для которого функция распределения принимает значение p ($0 < p < 1$) или ее значение изменяется скачком от меньшего p до превышающего p).

Размахом называется разность между наибольшим и наименьшим наблюдаемыми значениями случайной величины (формула 6).

$$R = Xmax – Xmin \quad (6)$$

Размахом пользуются как мерой рассеивания в эмпирических распределениях при малом числе наблюдений (когда $n \leq 10$).

1.2 Контрольные карты как метод контроля производственных процессов

Когда речь идет о регулярно протекающих процессах, остро становится вопрос поиска практических способов их контроля. Чем раньше обнаружится отклонение или сбой, тем легче и дешевле будет восстановить корректную работу.

Исследовательский отдел General Motors вывел правило десятикратных затрат. Согласно ему, расходы на исправление ошибки, найденной на следующей стадии производства, в 10 раз превышают затраты при обнаружении дефекта на стадии, где он был допущен. С каждым последующим этапом производства издержки растут в геометрической прогрессии.

На практике в организации параллельно работают тысячи процессов. Вручную такие объемы информации невозможно анализировать. Поэтому важно внедрить их автоматизированный мониторинг.

Контрольная карта Шухарта — это временной график изменения параметров процесса для осуществления статистического контроля его стабильности, который представляет из себя визуальный инструмент контроля за

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

технологическим, деловым или другим процессом для обнаружения и анализа его отклонений. Данный метод был впервые предложен американским учёным и консультантом в области управления Уолтером Шухартом.

Основная задача карт Шухарта — определить, возможно ли использовать последовательность данных для того, чтобы предсказать поведение процесса в будущем. В случае, если данные использовать нельзя, карты сигнализируют о его нестабильности, а если их применение возможно, являются основой для построения прогноза [20].

Контрольные карты Шухарта строятся для того, чтобы исключить специальные причины появления изменчивости. Тогда процесс становится стабильным и предсказуемым. Кроме того, повышается его качество и уменьшаются издержки.

По вертикальной оси графика откладываются множественные значения параметра процесса, а по горизонтальной — номера подгрупп, для которых актуальны данные значения.

Важно учитывать, что этот метод применим для регулярных (исполняемых через определенный промежуток времени) процессов, т.к. для построения нужно большое количество значений его параметра.

Пример контрольной карты представлен на рисунке 1.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

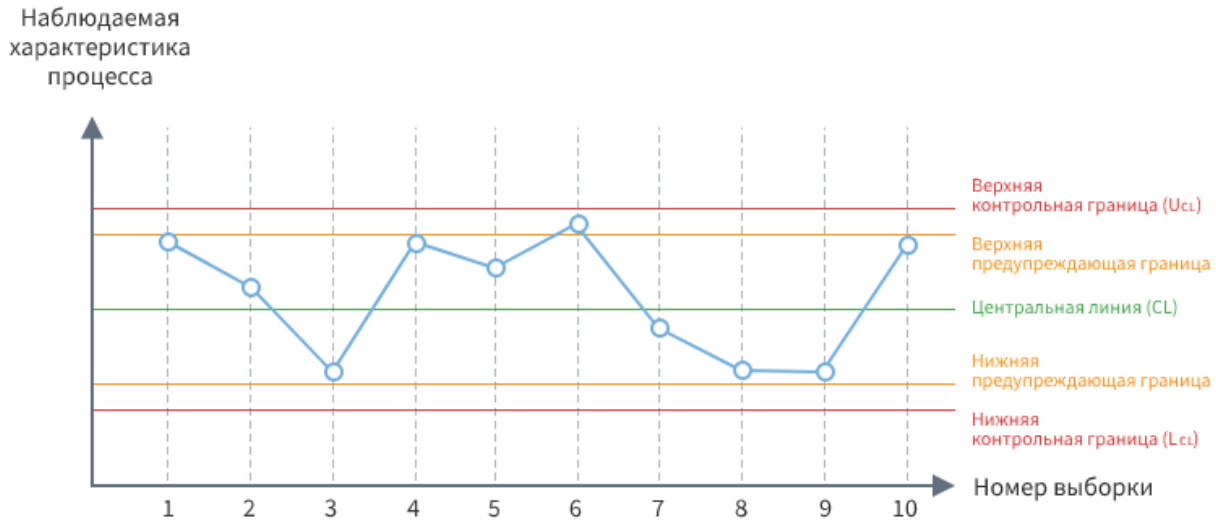


Рисунок 1 – Контрольная карта

На контрольную карту наносятся значения малых выборок изучаемой характеристики процесса, измеренных в порядке его протекания. Карты Шухарта имеют центральную линию, за основу которой чаще всего берётся среднее арифметическое значение показателя процесса. По обе стороны от нее на расстоянии 3σ откладывают верхнюю и нижнюю контрольные границы, где σ – их среднее квадратичное отклонение [21].

CL — центральная линия. Соответствует опорному значению характеристики. Центральная линия обычно определяется как среднее арифметическое значение. При этом стоит учитывать, что в зависимости от задачи управления это может быть также и намеченное (желаемое) целевое значение.

UCL и LCL — верхний и нижний контрольные пределы. Устанавливают значение коридора параметра процесса с обеих сторон в направлении выходной линии. Пока параметры не выходят за пределы, процесс считается стабильным с точки зрения статистики.

Важно, чтобы контрольная граница была установлена правильно. Если она расположена слишком близко к центральной линии, проблемы могут быть

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

обнаружены без каких-либо проблем. И наоборот, если она слишком широкая, источник определенных отклонений, может быть, не обнаружен.

Если распределение значений характеристик процесса является нормальным, 99,73% всех значений должны попадать в этот диапазон. Для процессов, находящихся в устойчивом состоянии, вероятность того, что точка на графике выйдет из-под контроля, составляет около 0,003. Если значение попадает в контрольный диапазон или выходит за его пределы, считается, что процесс вышел из устойчивого состояния [21].

Кроме того, обычно отмечаются "предупреждающие" границы с расстоянием ±2σ, указывающие на приближение значения процесса к верхней или нижней границе.

Использование контрольных карт помогает обеспечить стабильность процесса путем мониторинга изменений процесса в режиме реального времени.

Особенностью контрольных карт является то, что они нечувствительны к небольшим изменениям на уровне процесса. Это означает, что они реагируют только на аномальное поведение, которое приводит к нестабильному поведению процесса [22].

1.3 Виды контрольных карт

Контрольные карты классифицируются в зависимости от типа данных, которые в них отражаются:

1. Для количественных данных, особенно подходит для измерения конкретных результатов процесса.
2. Для альтернативных данных, такие карты характеризуются дискретными значениями процесса (да/нет).

В большинстве случаев на практике предпочтительнее первый тип диаграмм. Однако собирать данные с его помощью сложнее [23].

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Контрольные карты для количественных данных имеют следующие преимущества:

1. Большинство процессов имеют измеряемые характеристики.
2. Измеренные значения содержат больше информации, чем простое "да" или "нет".
3. Можно анализировать эти характеристики отдельно от установленных требований.

Значения параметров процесса для каждой контрольной карты могут быть

не заданы. В этом случае данные процесса анализируются на основе самих значений процесса. Данные процесса ищут отклонения, которые могут быть вызваны только неслучайными причинами.

Заданные значения параметров процесса обычно поступают из запросов или из анализа данных за определенный период времени. В этом случае наблюдаемые значения сравниваются с целевыми значениями.

Разница между этими двумя типами карт заключается в том, что последняя имеет дополнительные критерии для положения центральной линии и контрольных границ.

1.4 Построение контрольных карт

Когда контрольные карты создаются впервые, сам процесс часто оказывается статистически нестабильным. Поэтому необходимо довести его до управляемого состояния.

Создание карты — это двухэтапный процесс. Первым этапом является настройка параметров контрольной карты. Сбор и обработка большого количества данных для того, чтобы экспериментально понять центральную линию и границы процесса. На момент сбора данных неизвестно, находится ли процесс в статистически контролируемом состоянии, поэтому предполагается,

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

что это эксперимент. Необходимо потратить время на выявление источника той или иной причины изменчивости, пока не станет ясно, что на карте нет сигнала. Сделав это, можно убедиться, что процесс стабилизировался [24].

Вторым этапом создания карты является поддержка процесса в статистически контролируемом состоянии. Значения, полученные на первом этапе, становятся параметрами карты управления на следующем этапе. На этом этапе важно поддерживать процесс в контролируемом состоянии и быстро выявлять любые особые причины, влияющие на процесс.

Основные критерии для определения нестабильности процесса:

- значения параметров, превышающие контрольные пределы;
- семь или более точек, расположенных последовательно по одну сторону от центральной линии;
- семь последовательно уменьшающихся или увеличивающихся точек;
- точки со значительными неслучайными изменениями значений.

Примеры случаев нестабильности процесса, подпадающих под данные критерии, представлены на рисунке 2:

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

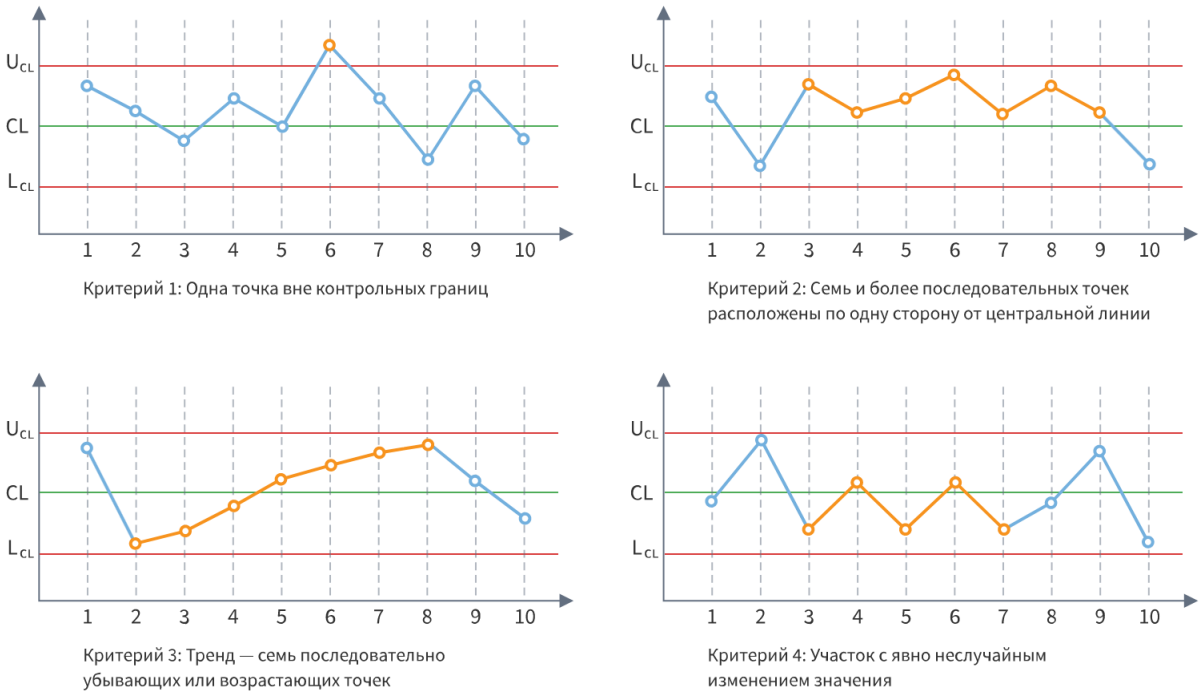


Рисунок 2 – Примеры случаев нестабильности процесса

Первоначально рекомендуется сосредоточиться на первом критерии оценки результатов — самом правиле Шухарта. Другие критерии могут быть добавлены по мере необходимости в зависимости от характера анализируемого процесса.

При долгосрочном анализе часто определяются собственные критерии выхода процесса из устойчивого состояния, которые также могут быть приняты во внимание. Однако следует помнить, что с каждым дополнительным критерием возрастает вероятность ложных тревог.

Поскольку процессы подвержены колебаниям, их необходимо отслеживать, а при отклонении от устойчивого состояния немедленно принимать управленческие решения.

Если процессы разовые, их можно отслеживать вручную. Если же это регулярный процесс, очевидно, что объем информации, необходимой для анализа, огромен, и анализ должен быть автоматизирован.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Например, банк ежедневно получает тысячи кредитных заявок. Хотя процесс принятия кредитных решений стандартизирован, все же существуют различия. С помощью исторических данных об эффективности можно рассчитать эталонный коридор и быстро определить точки отклонения от этого коридора [25].

На рисунке 3 показано, как этот процесс работает на практике. Точки хаотичны и часто выходят за пределы эталонного коридора.

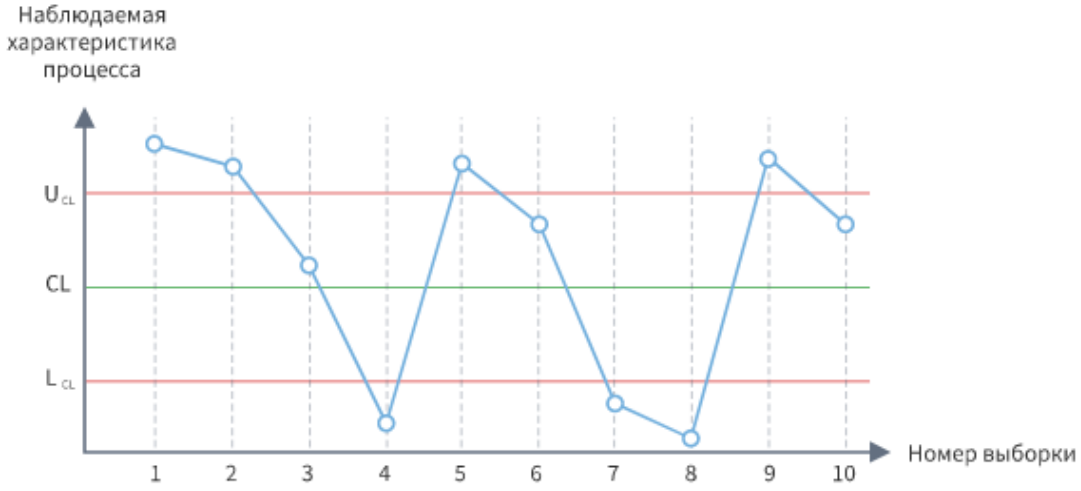


Рисунок 3 – Применение эталонного коридора в рабочему процессу

На приведенном ниже графике (рисунок 4) процесс перешел в контролируемое состояние. Все точки находятся в пределах контрольных границ, и структура формируется. Теперь можно прогнозировать его работу.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

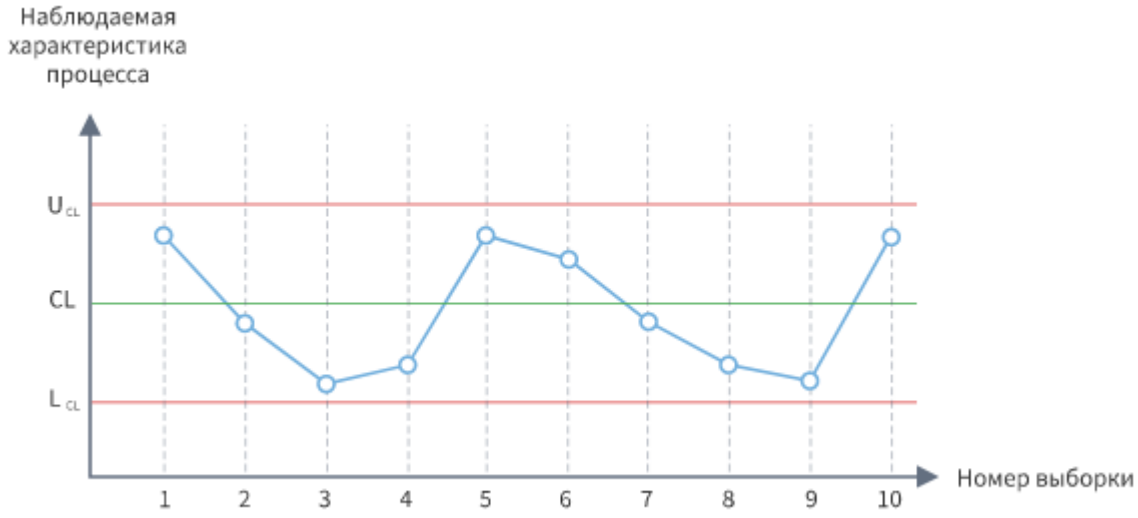


Рисунок 4 – Производственный процесс в контролируемом состоянии

Последний график (рисунок 5) показывает фазы управления, которые предотвращают переход состояния процесса в неустойчивое состояние. Точка 7 находится вне коридора управления и сигнализирует о необходимости применении управляющих воздействий.

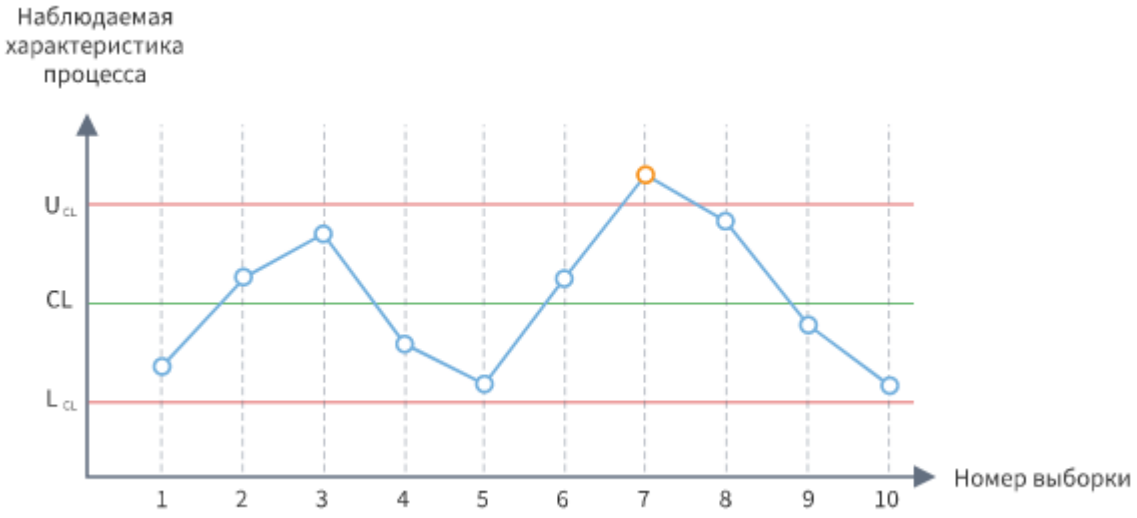


Рисунок 5 – Рабочий процесс, который может стать нестабильным

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Подводя итог, можно сделать вывод, что контрольные карты Шухарта являются универсальным инструментом для автоматизированного управления технологическими процессами. Стоит отметить следующие их преимущества:

1. Он позволяет быстро реагировать, когда процесс отклоняется от устойчивого состояния. Позволяет процессу быстро инициировать решение проблем.
2. Уменьшает вариации внутри процесса. Уменьшает общий разброс значений параметров после того, как процесс достиг устойчивого состояния.
3. Показывает тенденции в поведении процесса. Позволяет управлять процессом и прогнозировать его производительность [26].

1.5 Патентный анализ

На сегодняшний день применение статистических методов контроля производственных процессов являются актуальной темой для исследований. За последние годы было зарегистрировано большое количество патентов из разнообразных сфер применения, в которых используются рахличные имплементации статистических методов.

Мультивариативный статистический контроль процесса аддитивного производства с технологией лазерного плавления, Скотт Алан Голд. Данное изобретение представляет собой метод изготовления изделия при помощи лазерной 3D-печати, в котором применяются контрольные карты для визуализации стабильности состояния лазера. При обнаружении аномалий в параметрах лазера установка автоматически выключается. Данная работа интересна тем, что позволяет повысить стабильность применяемой технологии 3d-печати, известной непредсказуемостью результата и хода рабочего процесса.

Программа обнаружения аномалий, метод обнаружения аномалий и устройство обнаружения аномалий, Коу Маруяма. Данное изобретение является

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

решением для контроля качества за аппаратом, производящим полупроводники, с использованием контрольных карт и методов машинного обучения. Этот патент представляет интерес тем, что комбинирует несколько статистических методов не только для обнаружения аномалий, но и для своевременного реагирования на них.

Диагностика ошибок компонентов системы, Ян Панг. Рассматриваемый патент представляет из себя систему для мониторинга компьютерной сети и диагностики возникающих в ней ошибок при помощи контрольных карт и нейронных сетей. База данных с архивом предыдущих ошибок используется для сужения числа возможных причин ошибок и значительно ускоряет время диагностики сети. В данной работе статистические методы контроля производственных процессов используются для обнаружения аномалий и направлены на уменьшение простоек в работе системы.

Проанализировав большой объём научной литературы рассматривающей способы применения статистических методов контроля качества в производстве, можно сделать вывод о том, что использование данной технологии затрагивает множество сфер деятельности, а значит является популярным направлением развития и подтверждает актуальность настоящего исследования. В том числе изучение вышеперечисленного материала послужило основанием для формирования представления о работе с данными и поиском аномалий.

Основываясь на этом, можно сказать, что научная новизна моей работы заключается в использовании статистических методов контроля качества для контроля производственных процессов на предприятии с применением визуализации показателей датчиков оборудования при помощи контрольных карт для контроля стабильности рабочих процессов и своевременного обнаружения аномалий в работе с целью предотвращения простоев на производстве.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Также в ходе анализа можно заметить сравнительно малый объём отечественных патентов в этой сфере, что свидетельствует о малой степени разработанности данной темы в отечественной науке. Контроль каечстве производственных процессов – актуальная на сегодняшний день тема, и в текущих реалиях очень важно развивать отечественные технологии в данной области.

1.6 Результаты и выводы по первой главе

В рамках исследовательской работы были проанализированы существующие статистические методы контроля производственных процессов, а также ключевые параметры анализа производственного процесса с точки зрения процесса, такие как обычные и особые факторы, статистическая устойчивость производственного процесса. Также были изучены методы контроля качества продукции либо других результатов производственных процессов по количественному и альтернативному признакам.

Были проанализированы реализации статистических методов контроля производственных процессов, в том числе метод контрольных карт Шухарта. Были изучены виды контрольных карт для разных типов данных, алгоритм работы с контрольными картами, виды аномалий в производственных процессах и то, как они обнаруживаются на контрольных картах.

Анализ источников показал, что тема является актуальной на сегодняшний день. При этом можно отметить низкую степень разработанности темы в отечественной науке, что повышает актуальность данной работы с точки зрения импортозамещения.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

2 ВЫБОР СТЕКА ТЕХНОЛОГИЙ ДЛЯ РАЗРАБОТКИ

2.1 Язык программирования

Python — это высокоуровневый язык программирования общего назначения с динамической строгой типизацией и автоматическим управлением памятью, с акцентом на повышение производительности труда разработчиков, улучшение читаемости и качества кода, а также обеспечение переносимости написанных программ. Язык является полностью объектно-ориентированным, что означает, что все является объектом. Необычной особенностью языка является выделение блоков кода с отступами. Синтаксис основного языка минимален, и необходимость в справочной документации практически отсутствует. Сам язык известен как интерпретируемый язык и используется исключительно для написания сценариев. Недостатком языка является то, что он обычно медленнее и требует больше памяти, чем код, написанный на компилируемых языках, таких как C и C++.

Python — это мультипарадигмальный язык программирования, поддерживающий императивное, процедурное, структурное, объектно-ориентированное, метапрограммирование и функциональное программирование. Общие задачи программирования решаются с помощью динамической типизации [27].

Аспектно-ориентированное программирование частично поддерживается декораторами, а более полная поддержка обеспечивается дополнительными фреймворками. Такие методы, как контрактное и логическое программирование, могут быть реализованы с помощью библиотек и расширений.Основные архитектурные черты — динамическая типизация, автоматическое управление памятью, полная интроспекция, механизм обработки исключений, поддержка многопоточных вычислений с глобальной блокировкой интерпретатора (GIL), высокоуровневые структуры данных [28]. Данный язык поддерживает разбиение

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

программ на модули. Программные модули, в свою очередь, могут быть объединены в пакеты. Данный язык выбран в связи с наличием обширной экосистемы решений для проблем, связанных с наукой данных и статистическим анализом [29].

2.2 Среда программирования

Jupyter Notebook (Юпитер-ноутбук) — это программа-блокнот для записи, передачи и запуска кода. Ей можно пользоваться как своеобразной средой разработки. Она существует как веб-сервис, то есть доступна через интернет и позволяет передавать код другим разработчикам.

Чаще всего среду используют для Python, но она существует и для других языков программирования. Jupyter Notebook поддерживает языки Ruby, Perl, R, MATLAB, Julia и другие [30]. Часто это специализированные языки для задач, которые подразумевают быстрое написание и выполнение маленькой программы.

Отличие Юпитер-ноутбука от традиционных сред разработки — в его интерактивности. Программа позволяет запускать отдельные участки и блоки кода, выполнять их в любом порядке. А результаты работы сразу можно вывести в то же окно рядом с кодом. Пример графического интерфейса Jupyter Notebook с открытым блокнотом изображён на рисунке 6.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

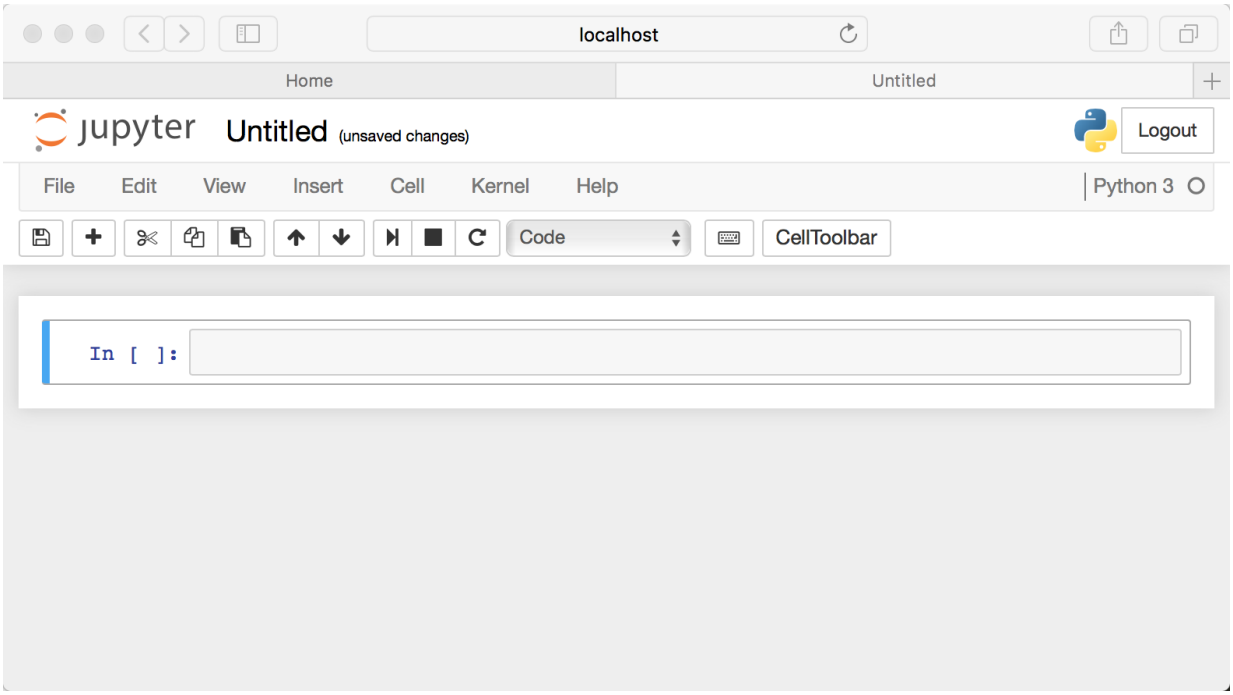


Рисунок 6 – Интерфейс Jupyter Notebook

Основные сферы использования среды — big data и data science, машинное обучение, математическая статистика и аналитика. В этих направлениях пригодилась способность Jupyter Notebook выводить данные в том же месте интерфейса, где написан код [31]. Получается, как бы страница блокнота: в одном месте собраны участки кода, результаты их выполнения, таблицы, участки и графики. Jupyter Notebook был выбран в качестве среды программирования как подходящее решение для проведения экспериментов и статистического анализа.

2.3 Менеджер пакетов

Anaconda — дистрибутив языков программирования Python и R. Он включает в себя набор популярных свободных библиотек, которые объединяет проблематика науки о данных и машинного обучения. Основная цель — поставка единым согласованным комплектом наиболее востребованных соответствующим кругом пользователей тематических модулей (таких как NumPy, SciPy, Astropy и

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

других) с разрешением возникающих зависимостей и конфликтов, которые неизбежны при одиночной установке. На момент 2019 года содержит более 1,5 тыс. модулей.

Главная особенность этого дистрибутива — оригинальный менеджер разрешения зависимостей conda с графическим интерфейсом Anaconda Navigator, который можно запускать без стандартного менеджера пакетов (например, pip в Python). Графический интерфейс Anaconda Navigator изображён на рисунке 7.

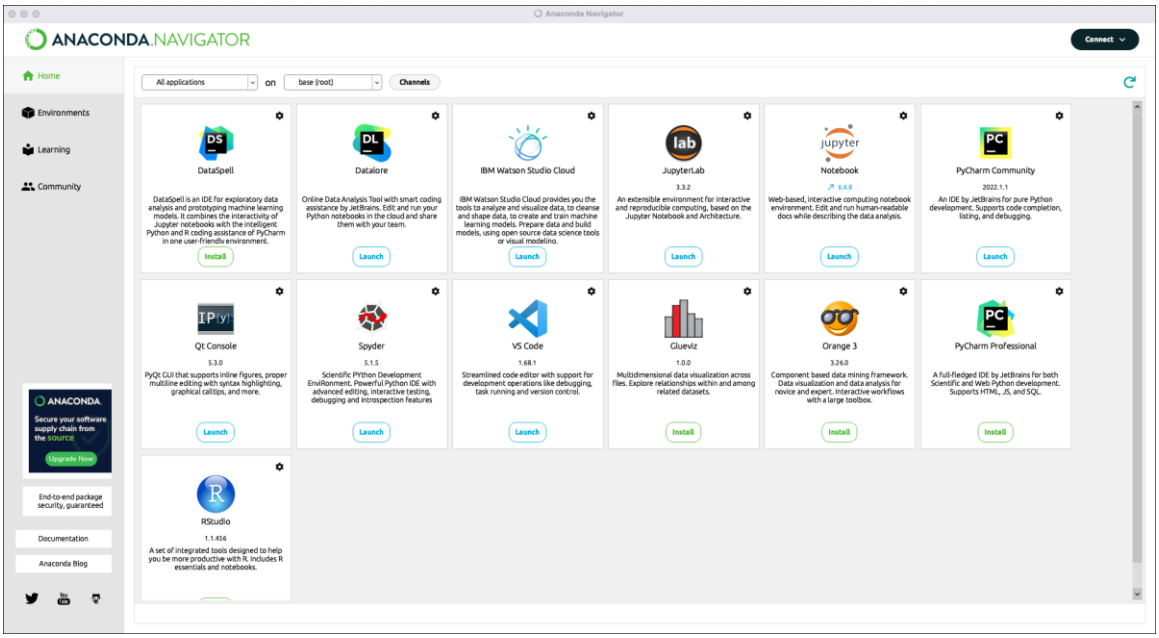


Рисунок 7 – интерфейс Anaconda Navigator

Дистрибутив можно загрузить один раз, а все последующие настройки, включая установку дополнительных модулей, можно выполнять в автономном режиме. Также можно поддерживать несколько независимых сред, разрешая зависимости версий по отдельности. Поддерживаются платформы Linux (x86-64), Windows (i686 и x86-64) и macOS. В качестве менеджера пакетов был выбрал Anaconda в связи с удобством относительно стандартных менеджеров пакетов Python, таких как pip, а также с тем, что этот менеджер пакетов создан преимущественно для анализа данных.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

2.4 Решение для соединения с базой данных

Для соединения с базой данных использовалась библиотека Psycopg2. Psycopg2 – одно из самых популярных решений для работы с базами данных PostrgreSQL. Эта библиотека включает в себя полную имплементацию спецификации Python DB API 2.0, которая представляет собой перечень стандартов для работы с базами данными на языке Python. Спецификация Python DB API 2.0 существует для максимально возможной унификации решений для работы с базами данных, что положительно сказывается на простоте понимания модулей, делает порт коды на другие базы данных намного проще, а также расширяет возможности работы с базами данных для Python.

Ещё одной отличительной чертой psycopg2 является потоковая безопасность. Концепция потоковой безопасности подразумевает, что код многопоточной программы исправно выполняет свои функции при выполнении его из нескольких потоков одновременно. Например, при доступе к разделяемым данным, потокобезопасный код должен обеспечивать правильный доступ для каждого потока. Потоковая безопасность достигается написанием кода в стиле, позволяющим частичное выполнение кода разными потоками. Части кода могут выполняться разными потоками по очереди. Данное свойство кода называется реентерабельностью. Реентерабельность требует сохранения состояния не в статических, либо же глобальных переменных, а в локальных, которые обычно находятся в стеке. Ещё одним свойством потокобезопасного кода является линеаризуемость, при которой код операций пишется так, что результат любого параллельного выполнения процедур эквивалентен их последовательному выполнению, то есть, функции не вмешиваются в работу других функций [32].

Psycopg2 используется для масштабных многопоточных приложений, которые создают и удаляют множество курсоров, областей памяти, которые содержат результат запроса SELECT. Psycopg2 позволяет производить с базой

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

данных больше число одновременных операций, которые изменяют хранимую в ней информацию, например, операций INSERT, которые добавляют новые заполненные значениями строки в таблицу, либо операций UPDATE, редактирующих имеющиеся в таблице записи.

Большая часть библиотеки Psycopg2 написана на языке C, и является обёрткой для библиотеки libpq, являющейся реализацией интерфейса для работы с базами данных PostgreSQL на языке C. Libpq является набором функцией, которые позволяют взаимодействовать с сервером PostgreSQL, обрабатывать запросы, и получать их результаты. Такой подход делает библиотеку psycopg2 надёжным и безопасным решением [33].

Библиотека psycopg2 позволяет работать с курсорами данных как на стороне клиента, так и на стороне сервера, а также поддерживает асинхронное взаимодействие с базой данных, при котором контакт между клиентом и сервером производится с временной задержкой. Psycopg2 также поддерживает асинхронные уведомления для сессий, в числе которых команды Listen и Notify.

Psycopg2 поддерживает использование команд COPY TO и COPY FROM, предоставляющим интерфейс для эффективного копирования данных из таблиц базы данных PostgreSQL в файлы и наоборот из файлов в таблицы базы данных PostgreSQL. Многие типы данных, которые используются в Python поддерживаются библиотекой «из коробки», и адаптированы под типы данных, использующиеся в PostgreSQL. Также в библиотеке присутствуют дополнительные расширения для адаптации различных типов данных. Psycopg2 поддерживает Юникод и синтаксис Python.

2.5 Решение для взаимодействия с базой данных

Для взаимодействия с базой данных и составления запросов использовалась библиотека SQLAlchemy. SQLAlchemy — это набор

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

инструментов Python SQL, который позволяет разработчикам получать доступ к базам данных SQL и управлять ими с помощью характерного для Python доменного языка. Вы можете написать запрос в виде строки или цепочки объектов Python для подобных запросов [34]. Работа с объектами предоставляет разработчикам гибкость и позволяет создавать высокопроизводительные приложения на основе SQL.

Базы данных SQL ведут себя тем менее как наборы объектов, чем большее значение начинают иметь размер и производительность; коллекции объектов ведут себя менее как таблицы и строки, чем больше значение абстракции начинает иметь значение. SQLAlchemy стремится учесть оба этих принципа.

SQLAlchemy считает базу данных механизмом реляционной алгебры, а не просто набором таблиц. Строки можно выбирать не только из таблиц, но также из объединений и других операторов выбора; любая из этих единиц может быть составлена в более крупную структуру. Язык выражений SQLAlchemy основан на этой концепции [35].

SQLAlchemy наиболее известен своим объектно-реляционным преобразователем (ORM), необязательным компонентом, который предоставляет шаблон преобразователя данных, где классы могут быть отображены в базу данных с открытым концом, несколькими способами, что позволяет объектной модели и схеме базы данных развиваться в чисто развязанный путь с самого начала.

Общий подход SQLAlchemy к этим проблемам полностью отличается от подхода большинства других инструментов SQL/ORM, основанного на так называемом подходе, ориентированном на комплементарность [36]; вместо того, чтобы скрывать детали SQL и объектно-реляционных отношений за стеной автоматизации, все процессы полностью раскрываются в наборе составных, прозрачных инструментов. Библиотека берет на себя работу по автоматизации

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

избыточных задач, в то время как разработчик сохраняет контроль над тем, как организована база данных и как построен SQL.

SQLAlchemy состоит из двух отдельных компонентов, известных как ядро и ORM. Ядро само по себе является полнофункциональным набором инструментов абстракции SQL, обеспечивающим плавный уровень абстракции для широкого спектра реализаций и вариантов поведения DBAPI, а также языка выражений SQL, который позволяет выражать язык SQL с помощью генеративных выражений Python. Система представления схемы, которая может как генерировать операторы DDL, так и анализировать существующие схемы, и система типов, которая позволяет любое сопоставление типов Python с типами базы данных, завершают систему. Затем Object Relational Mapper является необязательным пакетом, основанным на Core. Многие приложения построены строго на ядре, используя систему выражений SQL, чтобы обеспечить точный и точный контроль над взаимодействием с базой данных. Проще говоря, он позволяет пользователям подключать базы данных с помощью языка Python, выполнять SQL-запросы с помощью объектно-ориентированного программирования и оптимизировать рабочий процесс [37].

Более семи лет непрерывной разработки, профилирования и рефакторинга привели к созданию высокопроизводительного и точного инструментария, хорошо протестированного и развернутого в тысячах сред. Практически каждый основной компонент находится на второй или третьей полной итерации, SQLAlchemy 0.6 примерно в два раза быстрее, чем более старые версии 0.4, выпущенные всего несколько лет назад, а версии 0.7 и 0.8 продолжают улучшаться. Его необработанная скорость выполнения конкурентоспособна с сопоставимыми инструментами, а расширенные функции ORM, такие как его единица работы, коллекции в памяти, активная загрузка коллекций через соединения или вторичные подзапросы и другие оптимизации, позволяют ORM

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

SQLAlchemy выдавать меньше и более эффективные запросы, чем любая предыдущая версия [38].

Библиотека создана в соответствии с требованиями администраторов баз данных, включая возможность замены сгенерированного SQL операторами, оптимизированными вручную, полное использование параметров привязки для всех литеральных значений, полностью транзакционную и пакетную запись в базу данных с использованием шаблона Unit of Work. Все объектно-реляционные шаблоны разработаны с учетом использования надлежащей ссылочной целостности, и внешние ключи являются неотъемлемой частью ее использования.

SQLAlchemy придает большое значение тому, чтобы не мешать архитектуре базы данных и приложения. В отличие от многих инструментов, он никогда не «генерирует» схемы (не путать с выдачей определяемых пользователем DDL, в которых он превосходен) и не полагается на какие-либо соглашения об именах. SQLAlchemy поддерживает самое большое разнообразие баз данных и архитектурных проектов, насколько это возможно [39].

Система Unit Of Work, центральная часть Object Relational Mapper (ORM) SQLAlchemy, организует незавершенные операции вставки/обновления/удаления в очереди и сбрасывает их все одним пакетом. Для этого он выполняет топологическую «сортировку по зависимостям» всех измененных элементов в очереди, чтобы учитывать зависимости между строками, и группирует избыточные операторы вместе, где они иногда могут быть еще больше упакованы. Это обеспечивает максимальную эффективность и безопасность транзакций, а также сводит к минимуму вероятность взаимоблокировок. Создан на основе шаблона «Unit of work» Фаулера, а также Hibernate, ведущего объектно-реляционного преобразователя Java.

Создание запросов на основе функций позволяет создавать предложения SQL с помощью функций и выражений Python. Полный спектр возможностей

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

включает в себя логические выражения, операторы, функции, псевдонимы таблиц, выбираемые подзапросы, операторы вставки/обновления/удаления, коррелированные обновления, предложения select и EXISTS, предложения UNION, внутренние и внешние соединения, параметры связывания и свободное смешивание. буквальный текст внутри выражений. Сконструированные выражения компилируются для любого количества реализаций баз данных поставщиков (таких как PostgreSQL или Oracle), что определяется комбинацией «диалекта» и «компилятора», предоставляемых реализацией [40].

Различные части SQLAlchemy можно использовать независимо от остальных. Такие элементы, как пул соединений, компиляция операторов SQL и транзакционные службы, могут использоваться независимо друг от друга, а также могут быть расширены с помощью различных точек расширения. Интегрированная система событий позволяет вводить пользовательский код более чем в пятидесяти точках взаимодействия, в том числе в рамках выполнения основных операторов, генерации и самоанализа схемы, операций с пулом соединений, объектно-реляционной конфигурации, операций сохранения, событий изменения атрибутов и этапов транзакций. Новые элементы выражений SQL и пользовательские типы баз данных могут быть легко созданы и интегрированы.

ORM стандартизирует «декларативную» конфигурационную систему, которая позволяет создавать пользовательские классы, встроенные в метаданные таблицы, на которые они сопоставляются, так же, как и большинство других объектно-реляционных инструментов. Однако эта система совершенно необязательна — по своей сути ORM считает определяемый пользователем класс, связанные с ним метаданные таблицы и их сопоставление совершенно отдельными. С помощью функции mapper() любой произвольный класс Python можно сопоставить с таблицей или представлением базы данных.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Сопоставленные классы также сохраняют сериализуемость (pickling) для использования в различных системах кэширования.

ORM кэширует коллекции и ссылки между загруженными объектами, поэтому при каждом доступе не нужно выдавать SQL. Функция активной загрузки позволяет загружать целые графы объектов, связанных коллекциями и ссылками, с помощью нескольких или всего одного запроса, настраиваемого вплоть до точного количества операторов для каждого сопоставления или для каждого запроса, без каких-либо изменений в существующих запросах. требуется. Проблема "N+1", из-за которой ORM должна создавать отдельные операторы для всех объектов в коллекции, с SQLAlchemy осталась в прошлом.

В SQLAlchemy первичные и внешние ключи представлены наборами столбцов; действительно составное поведение реализовано с нуля. ORM имеет промышленную поддержку значимых (не суррогатных) первичных ключей, включая изменчивость и совместимость с ON UPDATE CASCADE, а также явную поддержку других распространенных составных шаблонов PK, таких как объекты «ассоциации» (отношения «многие ко многим» с дополнительное значение, придаваемое каждой ассоциации).

Самореферентные отображения поддерживаются ORM. Структуры списка смежности могут быть созданы, сохранены и удалены с помощью надлежащего каскадирования без накладных расходов на код, кроме тех, которые используются в структурах, не являющихся самореферентными. Загрузка самореферентных структур любой глубины может быть настроена на рекурсивную загрузку коллекций с помощью одного оператора с серией соединений (т. е. объединенной загрузки) или с помощью нескольких операторов, каждый из которых загружает полный набор записей на определенном уровне глубины (то есть подзапрос). Сохранение с таблицами, которые имеют взаимозависимые пары внешних ключей (т. е. «много х»/«один

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

конкретный x»), также изначально поддерживаются с помощью функции «после обновления» [41].

Доступна явная поддержка для наследования одной таблицы, конкретной таблицы и объединенной таблицы. Полиморфная загрузка (то есть запрос, который возвращает объекты нескольких типов-потомков) поддерживается для всех трех стилей. Загрузка каждого из них может быть оптимизирована таким образом, чтобы для полной загрузки полиморфного набора результатов использовалась только одна передача туда и обратно.

Средства объектно-реляционных запросов SQLA могут обрабатывать необработанные операторы SQL, а также простые наборы результатов, и экземпляры объектов могут генерироваться из этих результатов таким же образом, как и любая другая операция ORM. Любой сверхоптимизированный запрос, который вы или ваш администратор базы данных можете составить, вы можете запустить в SQLAlchemy, и пока он возвращает ожидаемые столбцы в наборе строк, вы можете получить из него свои объекты. Также можно использовать операторы, представляющие несколько типов объектов, с результатами, полученными в виде именованных кортежей, или с маршрутизацией зависимых объектов в коллекции родительских объектов.

Система типов допускает предварительную и последующую обработку данных как на уровне параметра привязки, так и на уровне набора результатов. Пользовательские типы можно свободно смешивать со встроенными типами. Доступны общие типы, а также типы, специфичные для SQL. SQLAlchemy также поддерживает Python 2.5 до последних версий 3.x. Другие поддерживаемые платформы включают Jython и Pypy. SQLAlchemy включает диалекты для SQLite, Postgresql, MySQL, Oracle, MS-SQL, Firebird, Sybase и других, большинство из которых поддерживает несколько DBAPI. Другие диалекты публикуются как внешние проекты. Для использования каждой конкретной базы

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

данных требуется соответствующая реализация DB-API 2.0 (или иногда одна из нескольких доступных).

Python DBAPI (аббревиатура от DataBase API) был создан, чтобы указать, как модули Python, которые интегрируются с базами данных, должны предоставлять свои интерфейсы. Хотя мы не будем взаимодействовать с этим API напрямую — мы будем использовать SQLAlchemy в качестве фасада — полезно знать, что он определяет, как должны вести себя общие функции, такие как соединение, закрытие, фиксация и откат. Следовательно, всякий раз, когда мы используем модуль Python, соответствующий спецификации, мы можем быть уверены, что найдем эти функции и что они будут вести себя так, как ожидалось. Одной из самых известных реализаций DBAPI для PostgreSQL является psycopg. Чтобы лучше понять спецификацию DBAPI, какие функции для нее требуются и как эти функции ведут себя, ознакомьтесь с предложением по улучшению Python, в котором она представлена [42].

Всякий раз, когда мы хотим использовать SQLAlchemy для взаимодействия с базой данных, нам нужно создать Engine. Механизмы в SQLAlchemy используются для управления двумя важными факторами: пулами и диалектами. В следующих двух разделах объясняется, что представляют собой эти две концепции, а сейчас достаточно сказать, что SQLAlchemy использует их для взаимодействия с функциями DBAPI. Чтобы создать движок и начать взаимодействовать с базами данных, нам нужно импортировать функцию create_engine из библиотеки sqlalchemy и вызвать ее. Например, функция create_engine('postgresql://usr:pass@localhost:5432/sqlalchemy') создает механизм PostgreSQL для связи с экземпляром, работающим локально на порту 5432 (по умолчанию). Он также определяет, что он будет использовать usr и передавать в качестве учетных данных для взаимодействия с базой данных sqlalchemy. Обратите внимание, что при создании механизма не происходит мгновенного подключения к базе данных. Этот процесс откладывается до того момента, когда

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

он необходим (например, когда мы отправляем запрос или когда создаем/обновляем строку в таблице).

Поскольку SQLAlchemy полагается на спецификацию DBAPI для взаимодействия с базами данных, поддерживаются наиболее распространенные доступные системы управления базами данных. PostgreSQL, MySQL, Oracle, Microsoft SQL Server и SQLite — все это примеры движков, которые мы можем использовать вместе с SQLAlchemy.

Пул соединений — одна из самых традиционных реализаций шаблона пула объектов. Пулы объектов используются как кэш предварительно инициализированных объектов, готовых к использованию. То есть вместо того, чтобы тратить время на создание часто используемых объектов (например, соединений с базами данных), программа извлекает существующий объект из пула, использует его по своему усмотрению и возвращает обратно, когда готово.

Основная причина, по которой программы используют преимущества этого шаблона проектирования, заключается в повышении производительности. В случае с соединениями с базами данных открытие и поддержание новых соединений требует больших затрат времени и ресурсов. Кроме того, этот шаблон упрощает управление количеством подключений, которые приложение может использовать одновременно.

В SQLAlchemy доступны различные реализации шаблона пула соединений. Например, создание Engine с помощью функции create_engine() обычно генерирует QueuePool. Этот тип пула настроен с некоторыми разумными значениями по умолчанию, такими как максимальный размер пула, равный 5 соединениям.

Поскольку обычно готовые к работе программы должны переопределять эти значения по умолчанию (чтобы точно настроить пулы в соответствии со своими потребностями), большинство различных реализаций пулов соединений

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

предоставляют аналогичный набор параметров конфигурации. В следующем списке показаны наиболее распространенные варианты с их описаниями:

- pool_size: устанавливает количество подключений, которые будет обрабатывать пул;
- max_overflow: указывает, сколько избыточных подключений (относительно pool_size) поддерживает пул;
- pool_recycle: настраивает максимальный возраст (в секундах) соединений в пуле;
- pool_timeout: определяет, сколько секунд программа будет ждать, прежде чем отказаться от получения соединения из пула.

2.5.1 SQLAlchemy.text

Для составления запросов к базе данных используется расширение «SQLAlchemy.text», позволяющий составлять запросы в базу данных в текстовом формате на языке выражений SQLAlchemy. Язык выражений SQLAlchemy представляет собой систему представления структур и выражений реляционных баз данных с использованием конструкций Python. Эти конструкции смоделированы так, чтобы максимально походить на конструкции базовой базы данных, обеспечивая при этом некоторую абстракцию различных различий реализации между серверными частями базы данных. Хотя конструкции пытаются представить эквивалентные концепции между серверными частями с согласованными структурами, они не скрывают полезных концепций, которые являются уникальными для определенных подмножеств серверных частей. Таким образом, язык выражений представляет собой метод написания выражений SQL, нейтральных к серверной части, но не пытается обеспечить, чтобы выражения были нейтральными к серверной части [43].

Язык выражений отличается от Object Relational Mapper, который представляет собой отдельный API, основанный на языке выражений. В то время

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

как ORM представляет собой высокоуровневый и абстрактный шаблон использования, который сам по себе является примером прикладного использования языка выражений, язык выражений представляет собой систему представления примитивных конструкций реляционной базы данных напрямую без мнения.

Хотя шаблоны использования ORM и языка выражений частично совпадают, сходство более поверхностно, чем может показаться на первый взгляд. Один из них подходит к структуре и содержанию данных с точки зрения определяемой пользователем модели предметной области, которая прозрачно сохраняется и обновляется из лежащей в ее основе модели хранения. Другой подходит к этому с точки зрения литеральной схемы и представлений выражений SQL, которые явно составлены в сообщения, потребляемые базой данных по отдельности.

Успешное приложение может быть создано с использованием исключительно языка выражений, хотя приложение должно будет определить свою собственную систему перевода концепций приложения в отдельные сообщения базы данных и из отдельных наборов результатов базы данных. В качестве альтернативы, приложение, созданное с помощью ORM, может в расширенных сценариях время от времени использовать язык выражений непосредственно в определенных областях, где требуются определенные взаимодействия с базой данных.

Переход от того, что понимается под текстовым выражением SQL, к конструкции Python, которая группирует компоненты вместе в программном стиле, может быть трудным. Вот почему SQLAlchemy позволяет вам просто использовать строки в тех случаях, когда SQL уже известен и нет сильной необходимости в том, чтобы оператор поддерживал динамические функции. Конструкция text() используется для составления текстового оператора, который передается в базу данных практически без изменений.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Конструкция text() также можно использовать для создания фрагментов SQL, которые могут свободно находиться внутри объекта select(), который принимает объекты text() в качестве аргумента для большинства своих функций построения. Ниже мы комбинируем использование text() с объектом select(). Конструкция select() обеспечивает «геометрию» оператора, а конструкция text() предоставляет текстовое содержимое этой формы. Мы можем построить оператор без необходимости ссылаться на какие-либо предварительно установленные метаданные таблицы. Хотя text() можно использовать в списке столбцов объекта select(), он имеет некоторые ограничения при составлении сгенерированного выбора, поскольку он не будет находится в коллекции SelectBase.selected_columns и будет исключен из коллекции подзапросов. В следующем разделе будет представлена конструкция literal_column(), которая является лучшим выбором для выражения имен отдельных столбцов в виде фрагментов SQL.

2.6 Инструмент для работы со статистическими функциями

Для работы со статистическими данными используется библиотека numpy. NumPy (Numerical Python) — это библиотека Python с открытым исходным кодом, которая используется практически во всех областях науки и техники. Это универсальный стандарт для работы с числовыми данными в Python, лежащий в основе научных экосистем Python и PyData. Пользователи NumPy включают всех, от начинающих программистов до опытных исследователей, занимающихся современными научными и промышленными исследованиями и разработками. NumPy API широко используется в Pandas, SciPy, Matplotlib, scikit-learn, scikit-image и большинстве других пакетов Python для обработки и анализа данных.

Библиотека NumPy содержит многомерные массивы и матричные структуры данных (вы найдете больше информации об этом в последующих разделах). Он предоставляет ndarray, объект однородного n-мерного массива, с

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

методами для эффективной работы с ним. Основное применение NumPy – математические операции над массивами. Он дополняет имеющийся набор инструментов Python структурами данных, которые позволяют самым эффективным образом проводить математические операции с массивами и матрицами, и предоставляет масштабную библиотеку математических функций продвинутого уровня, для работы с этими массивами и матрицами [44].

NumPy — основной пакет Python для научных вычислений. Это библиотека Python, которая предоставляет объекты многомерных массивов, различные производные объекты (такие как массивы масок и матрицы), а также набор подпрограмм для быстрого манипулирования массивами, включая математические, логические и графические операции, сортировку, выбор, ввод и вывод, дискретные преобразования Фурье, основы линейной алгебры, основные статистические операции и стохастическое моделирование. Это

В основе пакета NumPy лежит объект ndarray. Он инкапсулирует n-мерные массивы однородных типов данных и выполняет многие операции в скомпилированном коде для повышения производительности Есть несколько важных отличий между массивами NumPy и стандартными массивами Python:

1. Массивы NumPy, в отличие от списков Python (которые могут динамически расти), имеют фиксированный размер при создании: изменение размера массива создает новый массив и удаляет исходный.
2. Все элементы массива NumPy должны иметь одинаковый тип данных и, следовательно, иметь одинаковый размер в памяти. В качестве исключения можно иметь массив объектов (Python с NumPy). В этом случае у вас могут быть массивы с элементами разного размера.
3. Массивы NumPy позволяют легко выполнять сложные математические и другие операции над большими объемами данных. Такие операции часто могут быть выполнены более эффективно и с

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

меньшим количеством кода, чем при использовании встроенных массивов Python.

Все большее количество научных и математических пакетов на базе Python используют массивы NumPy. Эти пакеты обычно поддерживают ввод массивов Python, но часто преобразуют их в массивы NumPy перед обработкой и выводом. Другими словами, для эффективного использования большинства (возможно, большинства) современных научно-математических программ на базе Python недостаточно знать, как использовать встроенные в Python типы последовательностей; необходимо также знать, как использовать массивы NumPy.

Проблемы, связанные с размером и скоростью работы массивов, особенно важны в научных вычислениях. В качестве простого примера рассмотрим случай, когда каждый элемент одномерного массива умножается на соответствующий элемент другого массива той же длины. Если данные хранятся в двух списках a и b в Python, то для перебора каждого элемента можно использовать цикл for in range. Однако, если a и b содержат миллионы чисел, вы платите за неэффективность циклов Python: в C ту же задачу можно решить гораздо быстрее, написав простой цикл for (объявления и инициализация переменных, распределение памяти и т.д. для ясности игнорируются). Это экономит все накладные расходы, связанные с интерпретацией кода Python и манипулированием объектами Python, но за счет преимуществ, получаемых от кодирования на Python. Кроме того, требуемая работа по кодированию увеличивается с размерностью наших данных. Например, в случае двумерного массива код C (сокращенный, как и прежде) расширяется до цикла for внутри цикла for, и при дальнейшем увеличении измерений массива, число необходимых операций и кода будет только расти. В плане ndarrays, NumPy предлагает лучшее из двух миров. Операции с элементами выполняются в "режиме по умолчанию", но операции с элементами могут быть быстро выполнены с помощью

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

скомпилированного кода C; в NumPy операция "c = a * b" делает то же самое, что и предыдущий пример, с C В NumPy операция "c = a * b" делает то же самое, что и предыдущий пример. делает то же самое со скоростью, близкой к скорости языка C, но при этом код обладает простотой, ожидаемой от языка на основе Python. На самом деле, идиома NumPy еще проще. Этот последний пример иллюстрирует две функции NumPy, лежащие в основе большей части его возможностей: векторизацию и широковещание [45].

Векторизация описывает отсутствие каких-либо явных циклов, индексаций и т. д. в коде — эти вещи имеют место, разумеется, просто «за кадром» в оптимизированном, предварительно скомпилированном коде на C. Векторизованный код имеет множество преимуществ, среди которых:

- векторизованный код более лаконичен и удобен для чтения;
- меньше строк кода обычно означает меньше ошибок;
- код больше напоминает стандартную математическую нотацию (что, как правило, упрощает правильное кодирование математических конструкций);
- векторизация приводит к более «питоновскому» коду. Без векторизации наш код был бы замусорен неэффективными и трудными для чтения циклами for.

Бродкастинг — это термин, используемый для описания неявного поэлементного поведения операций [46]; вообще говоря, в NumPy все операции, не только арифметические, но и логические, побитовые, функциональные и т. д., ведут себя таким неявным поэлементным образом, т. е. транслируются. Более того, в приведенном выше примере a и b могут быть многомерными массивами одинаковой формы, или скаляром и массивом, или даже двумя массивами с разными формами, при условии, что меньший массив «расширяется» до формы большего таким образом, чтобы результат был однозначным [47]. Примеры бродкастинга массивов приведены на рисунке 8.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

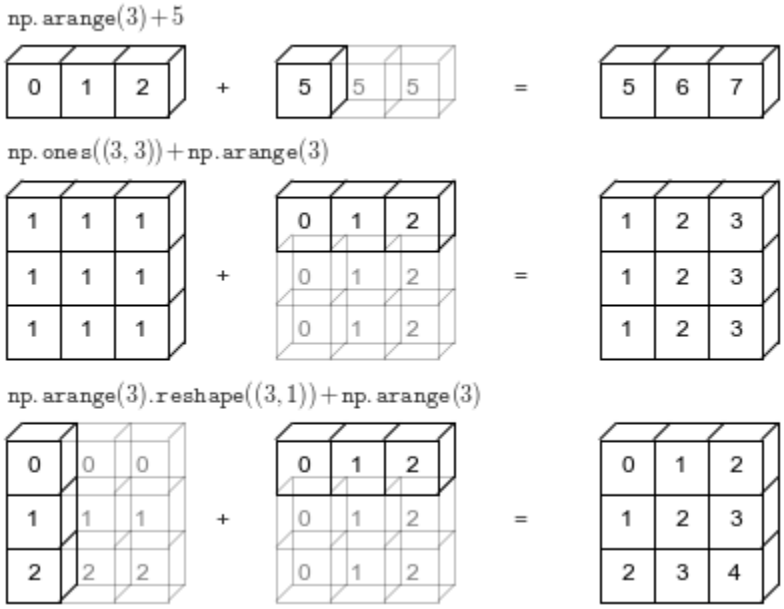


Рисунок 8 – Бродкастинг массива

NumPy полностью поддерживает объектно-ориентированный подход, опять же начиная с ndarray. Например, ndarray — это класс, обладающий многочисленными методами и атрибутами. Многие из его методов дублируются функциями в самом внешнем пространстве имен NumPy, что позволяет программисту кодировать в любой парадигме, которую он предпочитает. Эта гибкость позволила диалекту массива NumPy и классу ndarray NumPy стать де-факто языком многомерного обмена данными, используемым в Python.

NumPy предоставляет вам огромный выбор быстрых и эффективных способов создания массивов и манипулирования числовыми данными внутри них. Хотя список Python может содержать разные типы данных в одном списке, все элементы в массиве NumPy должны быть однородными. Математические операции, которые должны выполняться над массивами, были бы крайне неэффективны, если бы массивы не были однородными. Массивы NumPy быстрее и компактнее, чем списки Python. Массив потребляет меньше памяти и удобен в использовании. NumPy использует гораздо меньше памяти для хранения

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

данных и предоставляет механизм указания типов данных. Это позволяет еще больше оптимизировать код [48].

Массив — это центральная структура данных библиотеки NumPy. Массив представляет собой сетку значений и содержит информацию о необработанных данных, о том, как найти элемент и как его интерпретировать. Он имеет сетку элементов, которые можно индексировать различными способами. Все элементы имеют один и тот же тип, называемый массивом dtype.

Массив может быть проиндексирован кортежем неотрицательных целых чисел, логическими значениями, другим массивом или целыми числами. Ранг массива — это количество измерений. Форма массива представляет собой набор целых чисел, указывающих размер массива по каждому измерению.

Массив numpy иногда называют «ndarray», что является сокращением от «N-мерного массива». N-мерный массив — это просто массив с любым количеством измерений. Вы также можете услышать 1-D или одномерный массив, 2-D или двумерный массив и так далее. Класс NumPy ndarray используется для представления как матриц, так и векторов. Вектор — это массив с одним измерением (нет разницы между векторами-строками и векторами-столбцами), а матрица относится к массиву с двумя измерениями. Для трехмерных массивов или массивов более высокой размерности также обычно используется термин тензор [49].

Как и в других объектах-контейнерах Python, к содержимому массива можно получить доступ и изменить его путем индексации или нарезки массива. В отличие от обычных объектов-контейнеров, разные массивы могут совместно использовать одни и те же данные, поэтому изменения, сделанные в одном массиве, могут быть видны в другом.

Атрибуты массива отражают информацию, присущую самому массиву. Если вам нужно получить или даже установить свойства массива без создания нового массива, вы часто можете получить доступ к массиву через его атрибуты.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

NumPy также имеет довольно много полезных статистических функций, в том числе:

1. Функции numpy.amin() и numpy.amax(), которые возвращают минимум и максимум элементов заданного массива по указанной оси.
2. Функция numpy.ptp() возвращает диапазон (максимум-минимум) значений по оси.
3. Numpy.percentile, необходимая для вычисления процентиля. Процентиль (или центиль) — это мера, используемая в статистике, указывающая значение, ниже которого падает данный процент наблюдений в группе наблюдений. Функция numpy.percentile() принимает следующие аргументы: numpy.percentile(a, q, axis), в которой a – массив данных, q – номер рассчитываемого процентиля, axis – ось, по которой будет считаться процентиль.
4. Numpy.median, которая вычисляет медиану. Медиана определяется как значение, отделяющее верхнюю половину выборки данных от нижней половины.
5. Numpy.mean, вычисляющая среднее арифметическое. Среднее арифметическое — это сумма элементов по оси, деленная на количество элементов. Функция numpy.mean() возвращает среднее арифметическое элементов массива. Если указана ось, то он рассчитывается вдоль нее.
6. Функция numpy.average() вычисляет средневзвешенное значение элементов в массиве в соответствии с их соответствующим весом, заданным в другом массиве. Функция может иметь параметр оси. Если ось не указана, массив выравнивается. Средневзвешенное значение представляет собой среднее значение, полученное в

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

результате умножения каждого компонента на коэффициент, отражающий его важность.

7. Стандартное отклонение представляет собой квадратный корень из среднего значения квадратов отклонений от среднего значения. Формула стандартного отклонения выглядит следующим образом: std = sqrt (среднее (abs (x — x.mean ()) ** 2)). К примеру, если массив равен [1, 2, 3, 4], то его среднее значение равно 2,5. Следовательно, квадраты отклонений равны [2,25, 0,25, 0,25, 2,25], а квадратный корень из среднего значения, деленный на 4, т. е. sqrt (5/4) равен 1,1180339887498949.
8. Дисперсия — это среднее квадратов отклонений, т. е. среднее (abs (x — x.mean())**2). Другими словами, стандартное отклонение — это квадратный корень из дисперсии.

Почти каждый ученый, работающий с Python, использует возможности NumPy. NumPy составляет основу мощных библиотек машинного обучения, таких как scikit-learn и SciPy. По мере роста машинного обучения растет и список библиотек, построенных на NumPy. Возможности глубокого обучения TensorFlow имеют широкое применение — среди них распознавание речи и изображений, текстовые приложения, анализ временных рядов и обнаружение видео. PyTorch, еще одна библиотека глубокого обучения, популярна среди исследователей компьютерного зрения и обработки естественного языка. MXNet — еще один пакет ИИ, предоставляющий схемы и шаблоны для глубокого обучения.

Статистические методы, называемые ансамблевыми методами, такие как биннинг, бэггинг, суммирование и повышение, входят в число алгоритмов машинного обучения, реализуемых такими инструментами, как XGBoost, LightGBM и CatBoost — одними из самых быстрых механизмов логического вывода. Yellowbrick и Eli5 предлагают визуализации машинного обучения.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

NumPy является важным компонентом растущего ландшафта визуализации Python, который включает в себя Matplotlib, Seaborn, Plotly, Altair, Bokeh, Holoviz, Vispy, Napari и PyVista, и это лишь некоторые из них.

Ускоренная обработка больших массивов NumPy позволяет исследователям визуализировать наборы данных, намного большие, чем может обработать собственный Python.

2.7 Решение для работы с массивами данных

Для работы с датафреймами и массивами данных также использовалась библиотека Pandas. pandas — это пакет Python, который предоставляет быстрые, гибкие и выразительные структуры данных, предназначенные для того, чтобы сделать работу с «реляционными» или «помеченными» данными простой и интуитивно понятной. Он призван стать фундаментальным строительным блоком высокого уровня для практического анализа данных в реальном мире в Python. Кроме того, у него есть более широкая цель — стать самым мощным и гибким инструментом анализа/манипулирования данными с открытым исходным кодом, доступным на любом языке. Он уже на пути к этой цели. Основные функции библиотеки pandas включают, помимо прочего [50]:

- быстрый и эффективный объект DataFrame для манипулирования данными со встроенной индексацией;
- инструменты для чтения и записи данных между структурами данных в памяти и различными форматами: CSV и текстовые файлы, Microsoft Excel, базы данных SQL и быстрый формат HDF5;
- интеллектуальное выравнивание данных и интегрированная обработка отсутствующих данных: автоматическое выравнивание на основе меток в вычислениях и легкое преобразование беспорядочных данных в упорядоченную форму;
- гибкое изменение формы и поворот наборов данных;

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

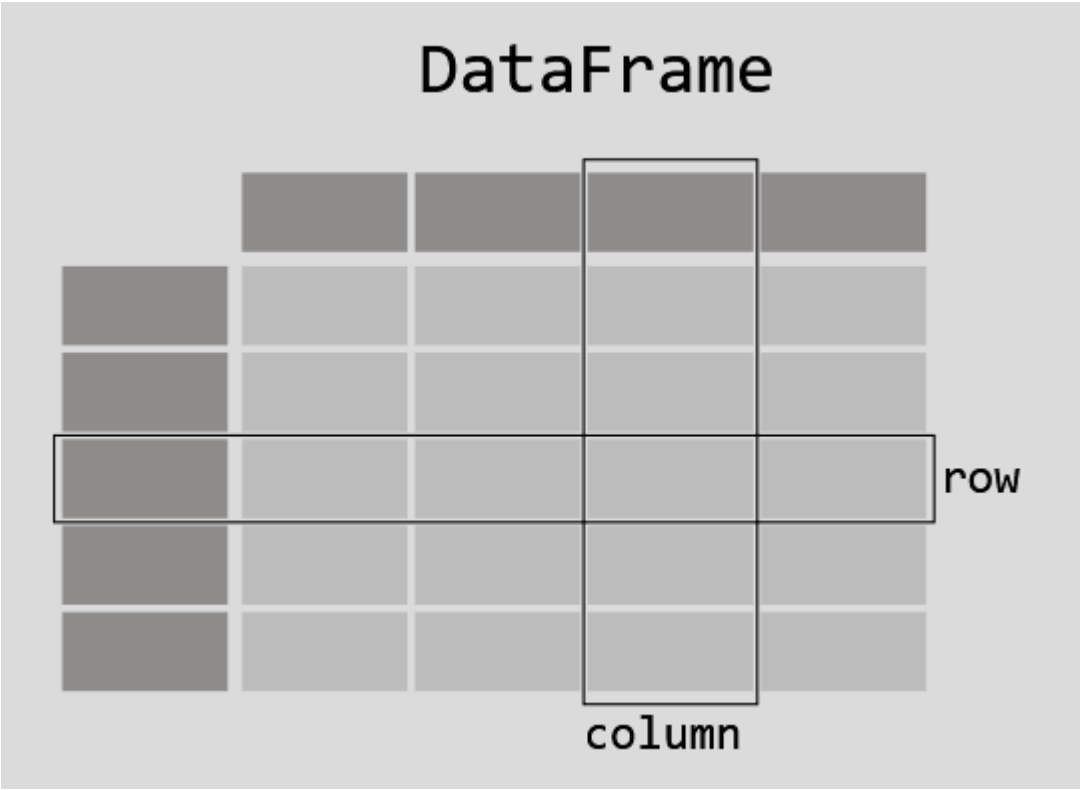
- интеллектуальная нарезка на основе меток, причудливая индексация и подмножество больших наборов данных;
- столбцы можно вставлять и удалять из структур данных для изменения размера;
- агрегирование или преобразование данных с помощью мощного механизма группировки, позволяющего выполнять операции «разделение-применение-объединение» над наборами данных;
- высокопроизводительное слияние и объединение наборов данных;
- иерархическая индексация осей обеспечивает интуитивно понятный способ работы с многомерными данными в низкоразмерной структуре данных;
- функциональность временных рядов: генерация диапазона дат и преобразование частоты, статистика скользящего окна, сдвиг даты и отставание. Даже создавайте смещения времени для конкретных доменов и объединяйте временные ряды без потери данных;
- высоко оптимизирован для производительности, с критическими путями кода, написанными на Cython или C;
- простая обработка отсутствующих данных (представленных как NaN, NA или NaT) в виде данных с плавающей запятой, а также данных без плавающей запятой;
- изменяемость размера: столбцы можно вставлять и удалять из DataFrame и объектов более высокого измерения;
- надежные инструменты ввода-вывода для загрузки данных из плоских файлов (CSV и с разделителями), файлов Excel, баз данных и сохранения/загрузки данных из сверхбыстрого формата HDF5;
- иерархическая маркировка осей (можно иметь несколько меток на тик).

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Все чаще пакеты создаются на основе библиотеки pandas для удовлетворения конкретных потребностей в подготовке, анализе и визуализации данных. Это обнадеживает, потому что это означает, что pandas не только помогает пользователям справляться с их задачами по работе с данными, но и предоставляют разработчикам лучшую отправную точку для создания мощных и более целенаправленных инструментов обработки данных. Создание библиотек, которые дополняют функциональность библиотеки pandas, также позволяет разработке pandas оставаться сосредоточенной на ее первоначальных требованиях. Библиотеки с использованием pandas используются во многих технологиях из разных сфер.

При работе с табличными данными, такими как данные, хранящиеся в электронных таблицах или базах данных, pandas — правильный инструмент для вас. Pandas позволяет пользователю исследовать, очищать и обрабатывать данные. В pandas таблица данных называется DataFrame. Пример структуры Датафрейма находится на рисунке 9.



Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Рисунок 9 – Структура датафрейма

DataFrame — это двумерная помеченная структура данных со столбцами потенциально разных типов. Вы можете думать об этом как об электронной таблице или таблице SQL или наборе объектов Series [51]. Как правило, это наиболее часто используемый объект в pandas. Как и Series, DataFrame принимает множество различных типов входных данных:

- dict одномерных ndarrays, списков, dicts или серий;
- 2-D numpy.ndarray;
- структурированный или записывающий ndarray;
- series;
- другой датафрейм.

Наряду с данными можно дополнительно передавать аргументы индекса (метки строк) и столбцов (метки столбцов). Если вы передаете индекс и/или столбцы, вы гарантируете индекс и/или столбцы результирующего DataFrame. Таким образом, dict of Series плюс определенный индекс отбросит все данные, не соответствующие переданному индексу. Если метки осей не переданы, они будут построены из входных данных на основе правил здравого смысла.

Dataframe может быть создан из dict of Series или dicts [52]. Результирующий индекс будет объединением индексов различных серий. Если есть какие-либо вложенные словари, они сначала будут преобразованы в Series. Если столбцы не переданы, столбцы будут упорядоченным списком ключей dict. Доступ к меткам строк и столбцов можно получить соответственно, обратившись к атрибутам индекса и столбцов [53].

В качестве инструмента для работы с массивами данных была выбрана библиотека pandas в связи с широким и гибким инструментарием для манипуляций со структурированными табличными данными.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

2.8 Решение для визуализации графиков

Для визуализации данных использовалась библиотека matplotlib. Matplotlib — это кроссплатформенная библиотека для визуализации данных и графического построения (гистограммы, точечные диаграммы, гистограммы и т. д.) для Python и его числового расширения NumPy. Таким образом, он предлагает жизнеспособную альтернативу MATLAB с открытым исходным кодом. Разработчики также могут использовать API-интерфейсы matplotlib (интерфейсы прикладного программирования) для встраивания графиков в приложения с графическим интерфейсом [54].

Сценарий Python matplotlib структурирован таким образом, что в большинстве случаев для создания графика визуальных данных достаточно нескольких строк кода. Слой сценариев matplotlib перекрывает два API:

1. API pyplot представляет собой иерархию объектов кода Python, возглавляемую matplotlib.pyplot.
2. Коллекция объектно-ориентированных API объектов, которые могут быть собраны с большей гибкостью, чем pyplot. Этот API обеспечивает прямой доступ к внутренним слоям Matplotlib.

API pyplot имеет удобный интерфейс с отслеживанием состояния в стиле MATLAB. На самом деле, Python-библиотека matplotlib изначально была написана как альтернатива MATLAB с открытым исходным кодом. OO API и его интерфейс более настраиваемы и мощны, чем pyplot, но считаются более сложными в использовании. В результате чаще используется интерфейс pyplot, и в этой статье он упоминается по умолчанию.

Понимание API pyplot matplotlib является ключом к пониманию того, как работать с графиками:

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

1. matplotlib.pyplot.figure: Figure — это контейнер верхнего уровня. Он включает в себя все, что визуализируется на графике, включая одну или несколько осей.
2. matplotlib.pyplot.axes: Оси содержат большинство элементов графика: Axis, Tick, Line2D, Text и т. д., а также задают координаты. Это область, в которой отображаются данные. Оси включают ось X, ось Y и, возможно, также ось Z.

Matplotlib и его зависимости можно загрузить в виде двоичного (предварительно скомпилированного) пакета из индекса пакетов Python (PyPI) и установить с помощью следующей команды: “python -m pip install matplotlib”.

Matplotlib также доступен в виде нескомпилированных исходных файлов на GitHub. Компиляция из исходного кода потребует, чтобы ваша локальная система имела соответствующий компилятор для вашей ОС, все зависимости, сценарии установки, файлы конфигурации и исправления. Это может привести к довольно сложной установке. В качестве альтернативы рассмотрите возможность использования платформы ActiveState для автоматического создания библиотеки matplotlib из исходного кода и ее упаковки для вашей ОС. Когда библиотека matplotlib используется для создания графика, создается пользовательский интерфейс (UI) и структура меню. Пользовательский интерфейс можно использовать для настройки графика, а также для панорамирования/масштабирования и переключения различных элементов [55].

На сегодняшний день наиболее распространенным способом создания фигуры является использование интерфейса pyplot. Как отмечено в интерфейсах приложений Matplotlib (API), интерфейс pyplot служит двум целям. Один из них — запустить Backend и следить за окнами графического интерфейса. Другой — это глобальное состояние для осей и исполнителей, которое позволяет использовать краткий API для построения графиков. В приведенном выше примере мы используем pyplot для первой цели и создаем объект Figure. В

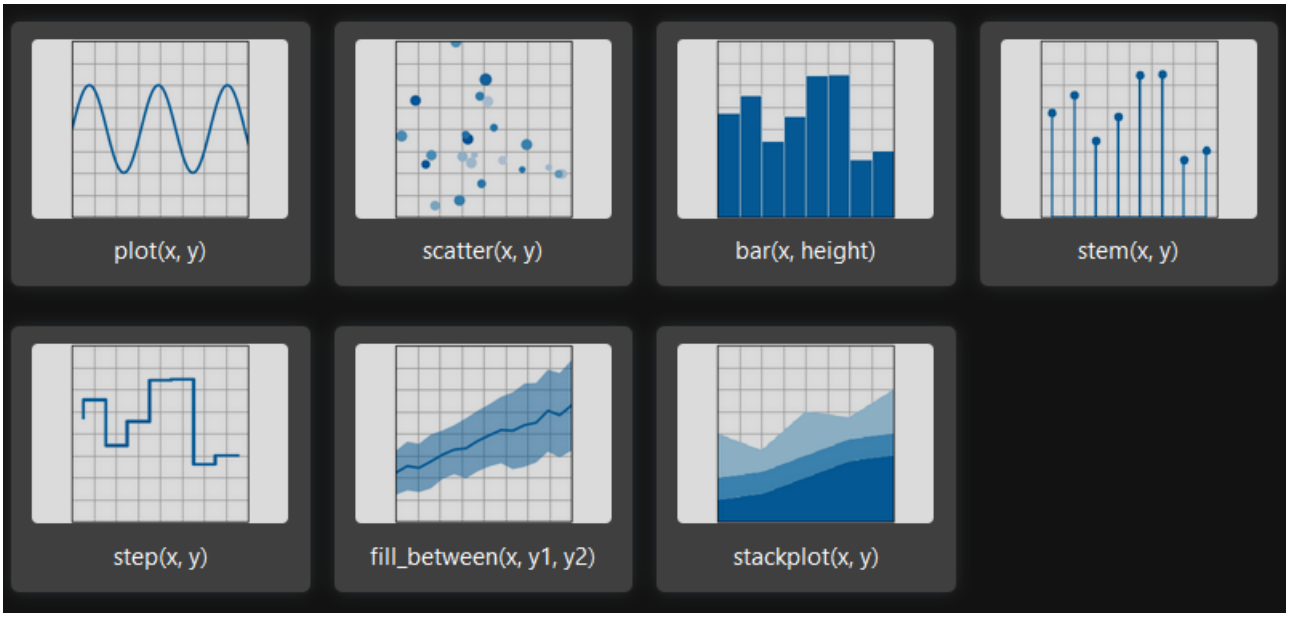
Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

качестве побочного эффекта fig также добавляется в глобальное состояние pyplot, и к нему можно получить доступ через gcf [56].

Пользователям обычно нужны оси или сетка осей при создании фигуры, поэтому в дополнение к фигуре существуют удобные методы, которые возвращают и фигуру, и некоторые оси. Простую сетку осей можно получить с помощью pyplot.subplots (который просто обертывает Figure.subplots). Иногда мы хотим иметь вложенную компоновку на рисунке с двумя или более наборами осей, которые не используют одну и ту же сетку подграфика. Мы можем использовать add_subfigure или subfigures для создания виртуальных фигур внутри родительской фигуры. Matplotlib позволяет строить следующие виды графиков:

- простейшие двухмерные графики, примеры графиков изображены на рисунке 10;
- графики полей и массивов;
- графики для статического анализа, такие как круги и гистограммы;
- графики неструктурированных координат;
- 3D-графики.



Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Рисунок 10 – Простейшие графики в matplotlib

В качестве решения для визуализации графиков была выбрана библиотека matplotlib в связи с простотой применения и соответствия доступного инструментария поставленным требованиям.

2.9 Решение для развёртывания формы

Основным продуктом данного проекта будет являться бэкенд и интерактивная веб-форма, позволяющая выбирать произвольные данные из предоставленной заказчиком БД. В данный момент существует множество решений для развёртывания веб-приложений для python.

Streamlit — это библиотека для преобразования скриптов Python в веб-приложения, доступные для совместного использования. Ориентированные на сообщество специалистов в области науки о данных и машинного обучения, приложения Streamlit позволяют создавать информационные панели и интерактивные ML-модели. Секрет популярности Streamlit заключается в ее простоте: использование библиотеки Streamlit быстро и интуитивно понятно, поэтому вы можете создавать приложения для науки о данных, не имея опыта веб-разработки. Для преобразования в интерактивный веб-интерфейс используется библиотека Python. Основные отличия заключаются в том, что Gradio используется для презентаций машинного обучения, а Streamlit — для создания приборных панелей; Plotly — это графическая библиотека, а Dash — это фреймворк с открытым исходным кодом для построения приложений обработки данных на Python, R и Julia. Это фреймворк с исходным кодом. (У Dash также есть корпоративная версия, но мы сосредоточимся на библиотеках с открытым исходным кодом). Panel — это библиотека Python с открытым исходным кодом для создания интерактивных веб-приложений приборных панелей, аналогичных

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

предыдущим трем вариантам. Panel очень гибкая и, как и Gradio, позволяет использовать любую графическую библиотеку, которая вам нравится. Однако, в отличие от Streamlit, его можно развернуть в Jupyter notebook. Panel Monitoring Panel также можно развернуть как отдельное веб-приложение, но, как и Plotly Dash, его можно развернуть самостоятельно. Anvil — это полноценная IDE для создания и развертывания полнофункциональных веб-приложений на базе Python. Вместо библиотеки Python, как в альтернативах, в Anvil есть конструктор страниц, облегчающий создание макетов и стилей пользовательского интерфейса. Поскольку Anvil — это не просто библиотека Python, он предоставляет интегрированную базу данных, аутентификацию пользователей и контроль версий, и может быть установлен в считанные секунды. В результате кривая обучения может быть немного длиннее, чем в других альтернативах [57].

Таким образом после анализа всех возможных опций в качестве решения для развёртывания формы была выбрана библиотека Streamlit как самая подходящая для выполнения поставленных задач.

2.10 Система управления базами данных

Производственные данные, с которыми будет производиться работа, хранятся в предоставленной заказчиком базе данных PostgreSQL. **PostgreSQL** — свободная объектно-реляционная система управления базами данных (СУБД). Для PostgreSQL существует множество реализация для платформ на основании Unix, а также Windows. PostgreSQL базируется на языке SQL и поддерживает многие из возможностей стандарта SQL:2011. При этом PostgreSQL имеет ряд своих ограничений:

- максимальный размер базы данных – 32 терабайта;
- максимум записей в таблице ограничен её размерами;
- максимальное число полей в записи в диапазоне от 250 до 1600, количество зависит от типов данных в полях;

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

- максимальный размер поля – 1 Гигабайт.

Функции — это блоки кода, которые выполняются на сервере, а не на клиенте базы данных. Они могут быть написаны на чистом SQL, но реализация дополнительной логики, такой как условные ветвления и циклы, выходит за рамки SQL и требует использования некоторых расширений языка. postgreSQL позволяет использовать функции, которые возвращают набор строк, которые можно использовать как обычный запрос Функции возвращают набор строк, которые можно использовать как обычный запрос. Функции могут выполняться с привилегиями автора и текущего пользователя. Функции иногда приравнивают к хранимым процедурам, но между ними есть разница. Начиная с версии 9 можно писать автономные блоки, которые позволяют выполнять код на языке программирования непосредственно на стороне клиента, без необходимости писать функции [58].

При этом PostgreSQL обладает рядом преимуществ перед другими решениями:

- высокая производительность и надежность механизмов транзакций и репликации;
- стандартная поддержка PL/pgSQL, PL/Perl, PL/Python и PL/Tcl; PL/Java, PL/PHP, PL/Py, PL/R, PL/Ruby, PL/Scheme, PL/sh и PL/V8 также доступны и могут быть загружены с расширениями C;
- наследование;
- возможность индексировать объекты геометрии (и особенно географии) и использовать расширения PostGIS, основанные на этих объектах;
- встроенная поддержка слабоструктурированных данных в формате JSON, которые можно индексировать;

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

- возможность создания новых типов данных, типов индексов, языков программирования и расширений, которые могут подключаться к любому внешнему источнику данных.

2.11 Графический клиент для СУБД

Работа с предоставленной базой данных осуществлялась в программе pgadmin. **PgAdmin** — это ведущий инструмент управления с открытым исходным кодом для Postgres, самой передовой в мире базы данных с открытым исходным кодом. PgAdmin 4 предназначен для удовлетворения потребностей как начинающих, так и опытных пользователей Postgres, предоставляя мощный графический интерфейс, который упрощает создание, обслуживание и использование объектов базы данных. Некоторые из элементов графического интерфейса представлены на рисунке 11.

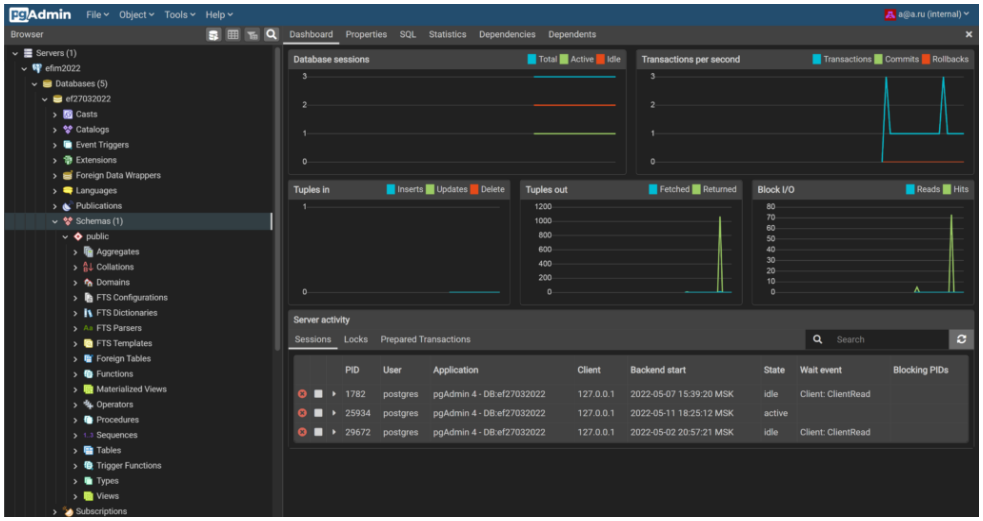


Рисунок 11 – Главный экран Pgadmin

2.12 Результаты и выводы по второй главе

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

В результате анализа требований заказчика, поставленных целей, а также изучения существующих решений для достижения поставленных задач был сформировал следующий стек технологий:

- язык программирования Python;
- среда разработки Jupyter Notebook;
- диспетчер пакетов Anaconda;
- библиотека psycopg2 для реализации связи с базой данных;
- библиотеке SQLAlchemy для составления запросов к базе данных;
- расширение библиотеки SQLAlchemy.text для более удобной работы с запросами;
- библиотека numpy для работы со статистическими функциями;
- библиотека pandas для манипуляций с датафреймами имеющихся данных;
- библиотека matplotlib для визуализации контрольных карт Шухарта в виде графиков.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

3 РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

3.1 Характеристика предприятия

Заказчиком проекта выступает компания «ДАТА-ЦЕНТР Автоматика». «ДАТА-ЦЕНТР Автоматика» — ведущий системный интегратор в области технической автоматизации. Компания является ведущим системный интегратор в области технической автоматизации. В число процессов, которыми занимается компания, входят автоматизация технологических процессов, управление производством, а также управление предприятием и промышленностью. Компания является ведущим системным интегратором в области автоматизации технических процессов, управления производством, предприятием и промышленностью.

На основе высокоинтеллектуальных и гибких решений компания разрабатывает интегрированные промышленные решения для интеграции производственных и управленческих процессов не только на отдельных предприятиях, но и в целых отраслях промышленности. Ее продукция основана на высокоинтеллектуальных системах автоматизации и виртуальном моделировании.

«ДАТА-ЦЕНТР Автоматика» обладает высокоспециализированной командой и уникальными технологиями. Инновационные разработки позволяют компании решать любые сложные задачи и удовлетворять потребности своих клиентов. «ДАТА-ЦЕНТР Автоматика» постоянно стремится к интеллектуальному и социальному развитию, что позволяет ей оставаться одним из ведущих разработчиков и поставщиков решений в области автоматизации и системной интеграции [59].

Компания разрабатывает передовые интеллектуальные системы, системы автоматизации процессов и системы управления производством в следующих областях:

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

- управление материальными потоками;
- расширенный анализ больших данных;
- базовая кафедра АБД и МВ;
- промышленная аналитика;
- планирование и управление железнодорожным транспортом;
- автоматизированные системы управления технологическими процессами — Системы управления технологическими процессами;
- компьютерное зрение — автоматизированные системы управления зрением;
- виртуальное моделирование;
- инновационные технологии.

Бизнес компании основан на интеграции технологий, что позволяет им разрабатывать и внедрять решения по автоматизации от одного технологического процесса на предприятии до целой отрасли, включая все уровни управления.

Достижения компании отмечены Правительством РФ, представлены на международных промышленных выставках, IP, инвестиционных и инновационных ярмарках, освещены в СМИ. Наиболее важные разработки компании запатентованы. «ДАТА-ЦЕНТР Автоматика» имеет зарегистрированный товарный знак и обладает всеми необходимыми лицензиями и сертификатами для работы над решениями и специальными требованиями, предъявляемыми к исполнителям [60].

Созданный в ходе данной исследовательской работы продукт будет интегрирован как один из компонентов более масштабной информационной системы. Конечный продукт — аналитическая система уровня предприятия. Система позволяет технологам и специалистам производить анализ производства с целью его оптимизации и улучшения. Конкретно компонент с контрольными картами позволит оперативно реагировать на отклонения в технологических

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

процессах. Соответственно при получении технологами оперативной информации по отклонениям позволит принять меры и избежать потенциальных будущих простоев в производстве, которые могут возникнуть в ответ на отклонения техпроцессов.

Во время разработки данного решения было решено использовать методологию бережливой разработки, входящей в семейство методологий Agile. Данная методология была выбрана в связи со следующими преимуществами:

- короткие итерации, что позволяет чаще коммуницировать с заказчиком для получения обратной связи, это будет важным преимуществом в связи с малыми сроками разработки;
- исключение потерь, подразумевающее минималистичный подход к функционалу, чтобы тратить время только на самое необходимое;
- целостное видение проекта, подразумевающее стандартизацию конечного продукта.

3.2 Анализ требований заказчика

Предполагаемый проект – программное обеспечение, которое будет являться инструментом контроля рабочих процессов. Решение должно выглядеть как интерактивная форма для выборки сигналов из предоставленной заказчиком базы данных.

Целевая аудитория проекта: работники завода, обращающиеся с показателями датчиков. Проект будет использоваться с рабочих компьютеров. Запуск проекта должен положительно повлиять на контроль за производственными процессами и выявление отклонений от нормы, что положительно может сказаться на выявлении и предотвращении ошибок, а также на непрерывности производственного процесса.

Программный продукт должен включать следующие функции:

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

- возможность выбора границ временного промежутка, от даты до часов и минут;
- возможность выбора типа сигналов, которые нужно демонстрировать;
- построение и визуализация контрольных карт Шухарта для каждого из выбранных производственных процессов.

Гайдлайны стиля – произвольные, форма должна представлять из себя несколько полей ввода для типа сигналов, начала и конца временного промежутка и пространство для демонстрации контрольных карт, данные и типы данных будут предоставлены заказчиком. Поскольку реальные данные защищены коммерческой тайной, вместо них были предоставлены сгенерированные. Сгенерированные данные будут обладать идентичной структурой и буду находиться в тех же численных значениях, что и реальные производственные.

3.3 Планирование

Следующим этапом процесса разработки является планирование функций ПО. Для полного понимания того, какой функционал должен быть реализован, был проведён эксплораторный анализ базы данных, предоставленных заказчиком. Поскольку реальные данные защищены коммерческой тайной, вместо них были предоставлены сгенерированные. Сгенерированные данные обладают идентичной структурой и будут находиться в тех же численных значениях, что и реальные производственные.

Предоставленная база данных состоит из трёх таблиц:

- таблица Signal, в которой расположены предполагаемые сигналы с датчиков на производстве;
- таблица Unit, в которой отмечены соответствия id сигналов и id производственных процессов, к которым они принадлежат;

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

- таблица Unit_type, в которой хранятся условные обозначения производственных процессов.

Таблица Signal состоит из трёх полей:

- поле Time, в котором хранится время регистрации сигнала;
- поле Signal_id, в котором хранится id сигнала в формате float;
- поле Signal_value, в котором хранится численное значение сигнала.

Примеры значений полей таблицы представлены ниже (рисунок 12):

	time	signal_id	signal_value
0	2023-05-03 02:54:06.717	1000777.0	325.0
1	2023-05-03 02:54:05.426	1000777.0	313.0
2	2023-05-03 02:54:05.965	1000777.0	318.0
3	2023-05-03 02:54:05.856	1000777.0	317.0
4	2023-05-03 02:54:05.533	1000777.0	314.0

Рисунок 12 – Пример строк из таблицы Signal

Таблица Unit состоит из четырёх полей:

- поле Id, в которой хранятся id сигналов;
- поле Unit_type_id, в которой хранятся айди производственных процессов, к каждому из которых принадлежит определённое число сигналов;
- поле Date_reg, в котором хранится дата регистрации данного процесса в системе;
- поле Etag, в котором хранится тег, уникальный для каждого производственного процесса. Данная переменная имеет служебное значение и в рамках данного проекта задействована не будет.

Примеры значений полей таблицы представлены ниже (рисунок 13):

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

	id	unit_type_id	data	date_reg	etag
0	15614670.0	1		14:30:25.262000	1684402225
1	15614671.0	1		14:30:25.262000	1684402225
2	15614672.0	1		14:30:25.262000	1684402225
3	15614673.0	1		14:30:25.262000	1684402225
4	15614674.0	1		14:30:25.262000	1684402225
...	...	...	...	...	...
1696	15406814.0	112		16:13:51.006000	1683890031
1697	15406815.0	112		16:13:51.006000	1683890031
1698	15406816.0	113		16:13:51.006000	1683890031
1699	15406817.0	115		16:13:51.006000	1683890031
1700	15406818.0	115		16:13:51.006000	1683890031

Рисунок 13 – Пример строк из таблицы Unit

Таблица unit_type состоит из четырёх полей:

- поле Id, в котором содержится айди производственного процесса;
- поле Type_name, в котором содержится название производственного процесса;
- поле Alias, в котором находится служебное обозначение производственного процесса;
- поле Metadata, в котором хранятся метаданные.

Последние два элемента имеют служебное значение и в рамках данного проекта использоваться не будут. Примеры значений полей таблицы представлены ниже (рисунок 14):

13	109	Конфигурация MTS.Состояние производства	stateProduction
14	2	ТестЕу2	test2
15	3	ТестЕу3	test3
16	114	Конфигурация MTS.Блок данных	dataBlock
17	115	Конфигурация MTS.Сигнал	signal

Рисунок 14 – Пример строк из таблицы unit_type

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

3.4 Дизайн

Как уже было упомянуто в первой главе, контрольные карты – это один из инструментов управления производственными процессами, который является достаточно гибким, чтобы найти множество различных применений. Контрольные карты имеют множество вариаций и подвидов в зависимости от рассчитываемых в них статистических величин. В широком смысле контрольные карты делятся на два подвида:

- контрольные карты для продолжительных данных, хорошо подходят для непрерывных данных, которые нужно отслеживать, хорошо подходит для отслеживания трендов и периодичности производственного процесса;
- контрольные карты для дискретных данных, больше подходят для поиска отдельных дефектов производственного процесса среди данных, не являющихся численными, либо не представленных непрерывным промежутком.

В связи со спецификой предоставленных заказчиком данных было решено использовать контрольные карты для продолжительных данных. В рамках проекта были рассмотрены следующие подвиды контрольных карт для продолжительных данных:

- карта, принимающая во внимание среднее значение и сдвиг размаха переменных;
- карта, принимающая во внимание среднее значение и размах переменных;
- карта, принимающая во внимание среднее значение и среднеквадратическое отклонение переменных.

Ввиду специфики данных, предоставленных заказчиком, был выбран тип контрольных карт, основанный на среднеквадратическом отклонении.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Для компонента веб-формы был разработан и утверждён с заказчиком макет интерфейса (рисунки 15-16).

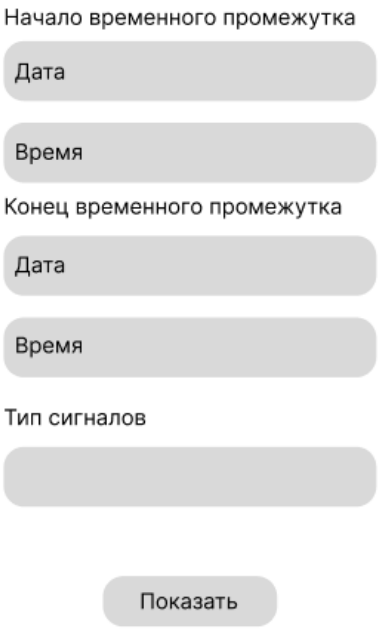


Рисунок 15 – Макет полей ввода

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

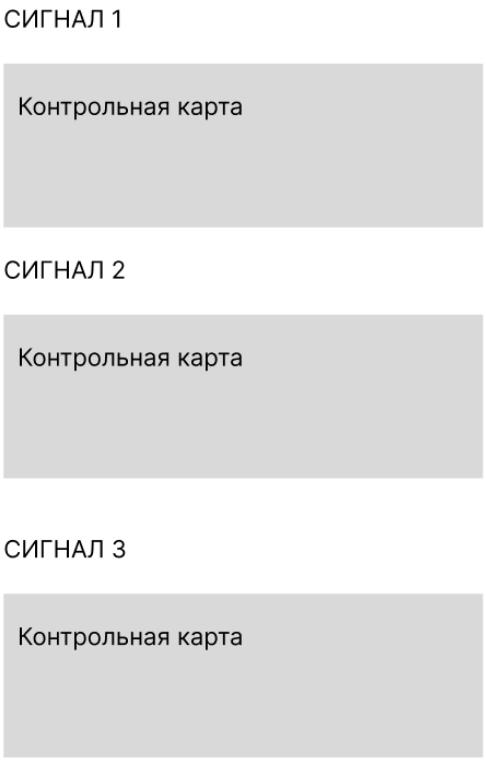


Рисунок 16 – Макет предполагаемого вывода

3.5 Разработка

Для построения контрольных карт был написан класс ShewhartControlModel, содержащий в себе несколько функций:

- функция Fit;
- функция Predict;
- функция Plot.

Для инициализации класса необходим только коэффициент k, определяющий границы контрольного промежутка. Данный коэффициент, помноженный на среднеквадратичное отклонение, определяет контрольные линии процесса, выход линии графика за которые будет означать наличие аномалий в производственном процессе.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

Функция fit принимает на вход массив данных, которые предстоит анализировать, среднее значение переменной и среднеквадратичное отклонение. При этом, если последние две величины не заданы пользователем, в функции реализованы методы вычисления данных величин. Для расчёта среднеквадратичного отклонения для числа переменных менее двадцати пяти также есть возможность воспользоваться табличными значениями коэффициентов для нахождения среднеквадратичного отклонения. Функция возвращает верхний и нижний контрольные пределы. Вызов данной функции необходим для нормальной работы других двух, поскольку в них используются вычисляемые здесь контрольные пределы процесса.

Функция predict принимает на вход массив данных, в котором нужно обнаружить исключения. Функция находит все значения, выбивающиеся за контрольные пределы, и возвращает массив, состоящий из порядкового номера точек-исключений и их значения.

Функция plot принимает на вход массив данных, для которых нужно построить график контрольной карты, а также dpi, определяющий масштабы графика.

В рамках проекта была разработана интерактивная форма выбора данных из базы, позволяющая пользователю выбирать произвольные временные промежутки и рабочие процессы. Интерактивная форма была развёрнута на локальной сети с помощью библиотеки streamlit. Имеющаяся форма имеет пять полей ввода:

- дата начальной границы временного промежутка;
- время начальной границы временного промежутка;
- дата конечной границы временного промежутка:
- время конечной границы временного промежутка;
- название производственного процесса в текстовой форме.

Графический интерфейс формы представлен ниже (рисунок 17).

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

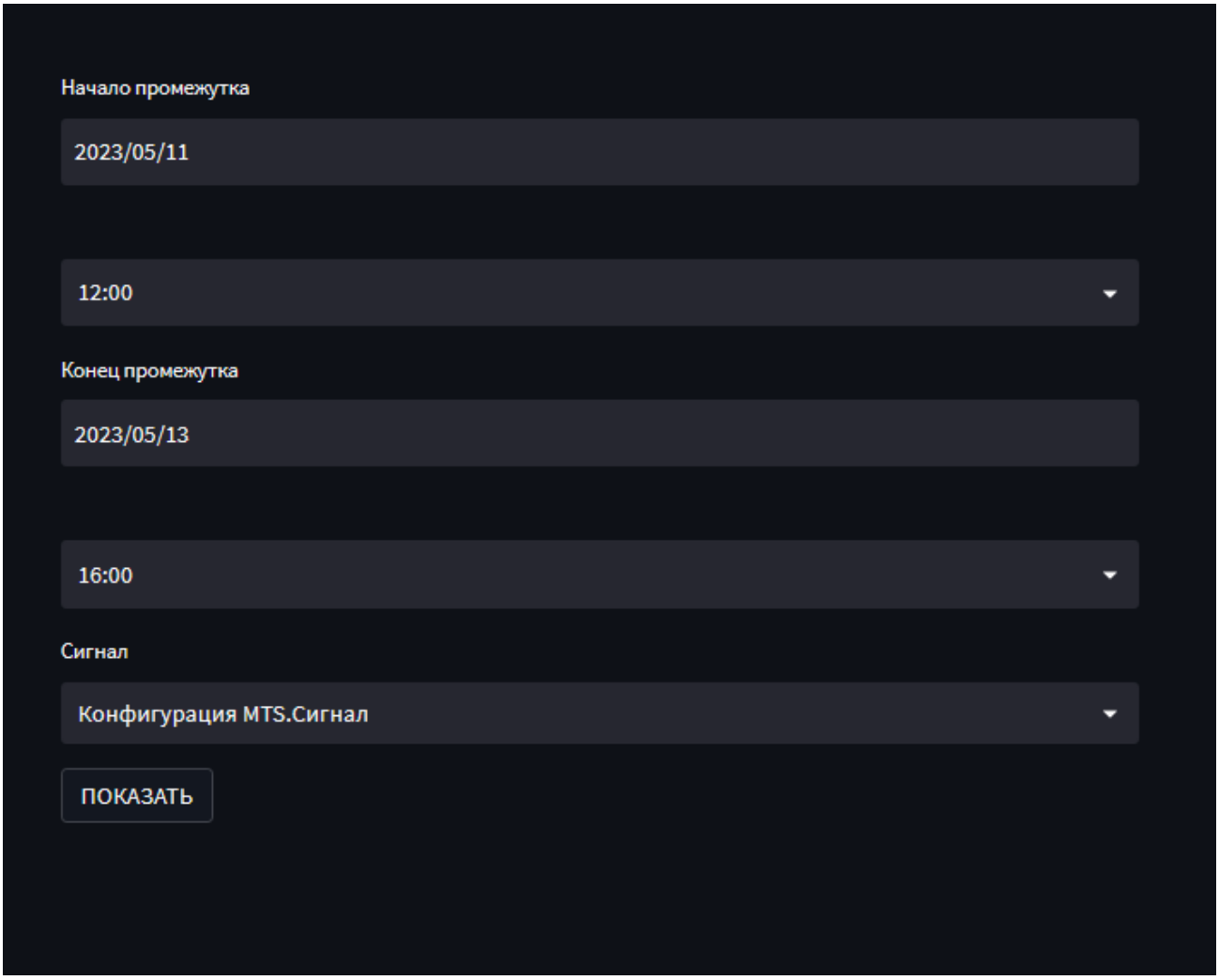


Рисунок 17 – Поля ввода

В качестве бэкенда для данной формы было написано несколько функций, позволяющих выбирать необходимые данные и строить контрольные карты Шухарта на выбранных промежутках:

- функция GetTimeProcessData принимает начало и конец временного промежутка в формате datetime, а также id необходимого сигнала, относящегося к производственному процессу, после чего возвращает массив значений сигналов, готовый к дальнейшему анализу;
- функция ShewhartChart принимает массив данных и с использованием класса ShewhartControlModel вычисляет

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

контрольные пределы для процесса, строит саму контрольную карту и возвращает массив значений, выходящих за контрольные пределы;

- функция GetIDbyProcess принимает название рабочего процесса и возвращает все id сигналов, относящиеся к выбранному процессу;
- функция AllChartsByProcess принимает название рабочего процесса и строит графики по каждому из относящихся к нему типов сигналов;
- функция ProcessIDfromTitle принимает текстовое название рабочего процесса и возвращает id данного процесса;
- функция GetTitleByID принимает id сигнала и возвращает id производственного процесса, к которому он принадлежит;
- функция GetAvailableTitleIDs принимает массив из id сигналов и возвращает массив всех id производственных процессов, к которым они принадлежат;
- функция GetAvailableTitles ищет в базе сигналов все уникальные id сигналов и возвращает массив названий всех производственных процессов, к которым они принадлежат.

3.6 Расчёт экономической эффективности

Поскольку данное решение будет являться компонентом более крупномасштабной информационной системы, расчёт экономической эффективности производится сразу для всей информационной системы, которая будет внедряться на предприятии. Компания «ДАТА-ЦЕНТР Автоматика» уже имеет опыт разработки и внедрения похожих информационных систем на производство, в частности внедрения статистических методов контроля качества на завод ЧерМК «Северсталь», где данные практики до этого не применялись.

Экономическим эффектом от внедрения информационной системы является снижение убытков из-за простоев в производстве, которые могут возникнуть из-за отклонений технологических процессов, поскольку данная

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

система позволит быстро обнаруживать и оперативно реагировать на любые отклонения. Расходы на разработку в месяц представлены в таблице 1.

Таблица 1 – Расходы на разработку ИС в месяц

Должность	З/п, руб.
Группа Интеграции с системами предприятия	
Middle-разработчик	96000
Middle-разработчик	96000
Team Lead	140000
Группа БД	
Middle-разработчик	96000
Middle-разработчик	96000
Team Lead	140000
Группа разработки статистических моделей	
Middle-разработчик	96000
Middle-разработчик	96000
Team Lead	140000
Группа визуализации	
Middle-разработчик	96000
Middle-разработчик	96000
Team Lead	140000
Итого:	1328000

При учёте того, что разработка и внедрение всей информационной системы целиком займёт один год, суммарные затраты на разработку и внедрение представлены в таблице 2.

Таблица 2 – Суммарные расходы на разработку ИС

Затраты в месяц, руб.	1328000
Срок разработки, мес.	12
Итого, руб.	15936000

Внедрённая на заводе ЧерМК информационная система позволила увеличить ежемесячную прибыль на 2250000 рублей. Таким образом можно рассчитать, что период возвращения инвестиций (ROI) для информационной системы составил 8 месяцев. Из этих расчётов можно получить подтверждение того, что внедрение цифрового решения для контроля качества производственных процессов с использованием статистических методов на

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

производство имеет реальный экономический эффект на его деятельность. Разрабатываемое в настоящее время решение будет внедряться на Абинском Электрометаллургическом заводе. По примерным подсчётам, от внедрения разрабатываемой информационной системы ожидается соизмеримый экономический эффект относительно затрат на разработку и полученной прибыли.

3.7 Результаты и выводы по третьей главе

В ходе исследовательской работы был произведён расчёт экономической эффективности проекта исходя из задач и проблем, которые заказчик хочет решить разработкой и внедрением этого проекта. На основе требований заказчика был сформулирован перечень функций, которыми конечный продукт должен обладать.

Была разработана интерактивная форма, позволяющая пользователю запрашивать данные сигналов, связанных с выбранным производственным процессом, на произвольном временном промежутке.

Для интерактивной формы был написан бэкенд с функциями построения и визуализации контрольных карт Шухарта для выбранных рабочих процессов.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

ЗАКЛЮЧЕНИЕ

В рамках исследовательской работы были проанализированы существующие статистические методы контроля производственных процессов, а также ключевые параметры анализа производственного процесса с точки зрения процесса, такие как обычные и особые факторы, статистическая устойчивость производственного процесса.

Также были изучены методы контроля качества продукции либо других результатов производственных процессов по количественному и альтернативному признакам.

Были проанализированы реализации статистических методов контроля производственных процессов, в том числе метод контрольных карт Шухарта. Были изучены виды контрольных карт для разных типов данных, алгоритм работы с контрольными картами, виды аномалий в производственных процессах и то, как они обнаруживаются на контрольных картах.

В результате анализа требований заказчика, поставленных целей, а также изучения существующих решений для достижения поставленных задач был сформировал следующий стек технологий:

- язык программирования Python;
- среда разработки Jupyter Notebook;
- диспетчер пакетов Anaconda;
- библиотека psycopg2 для реализации связи с базой данных;
- библиотеке SQLAlchemy для составления запросов к базе данных;
- расширение библиотеки SQLAlchemy.text для более удобной работы с запросами;
- библиотека numpy для работы со статистическими функциями;
- библиотека pandas для манипуляций с датафреймами имеющихся данных;

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

- библиотека matplotlib для визуализации контрольных карт Шухарта в виде графиков.

В ходе исследовательской работы был произведён расчёт экономической эффективности проекта исходя из задач и проблем, которые заказчик хочет решить разработкой и внедрением этого проекта. На основе требований заказчика был сформулирован перечень функционала, которым конечный продукт должен обладать.

Была разработана интерактивная форма, позволяющая пользователю запрашивать данные сигналов, связанных с выбранным производственным процессом, на произвольном временном промежутке.

Для интерактивной формы был написан бэкенд с функциями построения и визуализации контрольных карт Шухарта для выбранных рабочих процессов.

Подводя итог работы, можно сказать следующее: цель работы «создание программного продукта, который мог бы использоваться для контроля производственных процессов с использованием статистических методов» достигнута, выполнены следующие поставленные в начале работы задачи:

1. Изучены статистические методы контроля производственных процессов.
2. Изучены существующие реализации методов контроля производственных процессов.
3. Сформулированы требования к разрабатываемому инструменту
4. Сформирован стек технологий.
5. Создана интерактивная форма выбора данных из предоставленной заказчиком базы данных.
6. Разработан функционал для визуализации статистических показателей процесса.

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Studopedia. Статистические показатели возможностей процесса. Индекс пригодности и индекс воспроизводимости [Сайт] – URL: https://studopedia.ru/11_214820_stabilnost-tehnologicheskih-protsessov-osobie-i-obichnie-prichini-izmenchivosti-protsessa.html (Дата обращения — 03.05.2023)

2. Воеводкина Л.Г. Макростатистический инструментарий распределенных вычислительных систем [Текст]: Учеб. Пособие. – Изд-во Тамб. гос. техн. ун-та, 2007 – 80 с. – (Высшее образование)

3. ГОСТ 15895-77. «Статистические методы управления качеством продукции. Термины и определения»

4. ГОСТ Р 50779. «Статистические методы. Показатели возможностей процессов. Основные методы расчета»

5. Сизова Т.М. Статистика: Учебное пособие [Текст]: Учеб. Пособие. – СПб ГУИТМО, 2005. – 80 с. (Высшее образование)

6. Чернова Т.В. Экономическая статистика: Учебное пособие[Текст]: Учеб. Пособие. – Изд-во ТРТУ, 1999. 140 с. (Высшее образование)

7. Studfiles. Характеристики распределения случайных величин [Сайт] – URL: https://studfile.net/preview/4295847/page:5/ (Дата обращения — 03.05.2023)

8. Studopedia. Статистические показатели возможностей процесса [Сайт] – URL: https://studopedia.ru/11_19589_statisticheskie-pokazateli-vozmozhnostey-protsessa.html (Дата обращения — 03.05.2023)

9. Орлов А.И. Математика случая: Вероятность и статистика – основные факты [Текст]: Учеб. Пособие. – М.: МЗ-Пресс, 2004. – 110 с. (Высшее образование)

10. Allbest. Анализ рядов распределения. [Сайт] – URL: https://knowledge.allbest.ru/emodel/2c0a65625b3bc68b4c53b88521306d37_0.html (Дата обращения – 05.05.2023)

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

11. Бородачёв С.М. Статистические методы в управлении качеством [Текст]: Учеб. Пособие – Изд-во Урал. ун-та, 2016. — 87с.

12. Мойзес Б.Б. Статистические методы контроля качества и обработка экспериментальных данных. [Текст]: Учеб. Пособие — Изд-во Томского политехнического университета, 2016. — 119с.

13. Горяинова Е.Р., Панков А.Р., Платонов Е.Н. Прикладные методы анализа статистических данных. [Текст]: Учеб. пособие — Изд. дом Высшей школы экономики, 2012. — 310с

14. Хабр. Чёрный ящик статистического анализа. [Сайт] — https://habr.com/ru/companies/stepic/articles/250527/ (Дата обращения — 06.05.2023)

15. Studfile. Метод скользящей средней. [Сайт] — https://studfile.net/preview/1091716/page:2/ (Дата обращения — 06.05.2023)

16. ГОСТ 16504-81. «Система государственных испытаний продукции. Испытания и контроль качества продукции. Основные термины и определения.»

17. Adeptik. Контроль качества: что это такое, задачи. [Сайт] — https://adeptik.com/blog/kontrol-kachestva/ (Дата обращения — 11.05.2023)

18. Qualicont. Контроль качества продукции и работы предприятия. [Сайт] — https://www.qualicont.ru/quality/ (Дата обращения — 11.05.2023)

19. Bizlog. Общие функции управления качеством продукции. [Сайт] — http://bizlog.ru/lib/b2/4_4_1.htm (Дата обращения — 11.05.2023)

20. ГОСТ Р 50779.42-99 «Статистические методы. Контрольные карты Шухарта».

21. Diplomba. Построение контрольных карт в статистике [Сайт] – URL: https://diplomba.ru/work/73574 (Дата обращения — 03.05.2023)

22. Loginom. Что такое контрольные карты Шухарта [Сайт] – URL: https://loginom.ru/blog/shewhart-control-shart#control-charts (Дата обращения — 03.05.2023)

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

23. НГТУ НЭТИ. Контрольные карты Шухарта. [Сайт] – URL: https://ami.nstu.ru/~headrd/seminar/Kontrol_Q/Control_card.htm (Дата обращения – 05.05.2023)

24. Linco Platform. Контрольные карты Шухарта для внутрилабораторного контроля качества. [Сайт] https://lincoplatform.ru/Blog/C?N=127 (Дата обращения — 07.05.2023)

25. Six Sigma Study. What are X Bar R Control Charts? [Сайт] – URL: https://sixsigmastudyguide.com/x-bar-r-control-charts/ (Дата обращения – 04.05.2023)

26. Six Sigma Study. What are X Bar S Control Charts? [Сайт] – URL: https://sixsigmastudyguide.com/x-bar-s-chart/ (Дата обращения – 04.05.2023)

27. Interest Programs. История языка Python. [Сайт] – URL: https://www.interestprograms.ru/article-python-istoriya-sozdaniya-yazyka-programmirovaniya (Дата обращения – 05.05.2023)

28. Репозиторий БГУИР. Применение языка программирования Python в медицине. [Сайт] – URL: https://libeldoc.bsuir.by/bitstream/123456789/41809/1/1Feklenkova_Primeneniye.pdf (Дата обращения – 05.05.2023)

29. Академия информационных систем. Программирование на языке Python. Data Science. [Сайт] – URL: https://www.infosystems.ru/courses/python/PH007/ (Дата обращения – 06.05.2023)

30. Код. Что такое jupyter-ноутбук и зачем он нужен. [Сайт] – URL: https://thecode.media/jupyter/ (Дата обращения – 04.05.2023)

31. Машинное обучение. Jupyter Notebook. [Сайт] – URL: https://www.dmitrymakarov.ru/python/jupyter-notebook-14/ (Дата обращения – 05.05.2023)

32. Psycopg. Psycopg – PostgreSQL database adapter for Python [Сайт] – URL: https://www.psycopg.org/docs/ (Дата обращения – 04.05.2023)

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

33. Psycopg. Basic modular usage [Сайт] – URL: https://www.psycopg.org/docs/usage.html (Дата обращения – 04.05.2023)

34. GormAnalysis. Intro to SQLAlchemy. [Сайт] — https://www.gormanalysis.com/blog/intro-to-sqlalchemy/ (Дата обращения — 07.05.2023)

35. Altexsoft. Understanding object – relational mapping: pros, cons and types. [Сайт] — https://www.altexsoft.com/blog/object-relational-mapping/ (Дата обращения — 07.05.2023)

36. Dev. Introduction to object-relational mapping. [Сайт] — https://dev.to/tinazhouhui/introduction-to-object-relational-mapping-the-what-why-when-and-how-of-orm-nb2 (Дата обращения — 11.05.2023)

37. Python Central. Python’s SQLAlchemy vs Other ORMs [Сайт] – URL: https://www.pythoncentral.io/sqlalchemy-vs-orms/ (Дата обращения – 04.05.2023)

38. Хабр. Крадущийся тигр, затаившийся SQLAlchemy. [Сайт] – URL: https://habr.com/ru/articles/470285/ (Дата обращения – 04.05.2023)

39. Метанит. Основные операции с данными в SQLAlchemy. [Сайт] – URL: https://metanit.com/python/database/3.3.php (Дата обращения – 04.05.2023)

40. Tutorials Point. SQLAlchemy Core — Using Textual SQL. [Сайт] – URL: https://www.tutorialspoint.com/sqlalchemy/sqlalchemy_core_using_textual_sql.htm (Дата обращения – 05.05.2023)

41. SQL Alchemy. SQL Alchemy documentation. [Сайт] – URL: https://docs.sqlalchemy.org/en/14/core/tutorial.html (Дата обращения – 05.05.2023)

42. Хабр. Работа с базой данных: используем DB-API. [Сайт] – URL: https://habr.com/ru/articles/321510/ (Дата обращения – 06.05.2023)

43. Python Enhancement Proposals – Python Database API Specification v 2.0. [Сайт] – URL: https://peps.python.org/pep-0249/ (Дата обращения – 06.05.2023)

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

44. CoderLessons. NumPy – Статистические функции. [Сайт] – URL: https://coderlessons.com/tutorials/python-technologies/uchitsia-numpy/numpy-statisticheskie-funktsii (Дата обращения – 06.05.2023)

45. Real Python. Python statistics fundamentals. [Сайт] – https://realpython.com/python-statistics/ (Дата обращения – 06.05.2023)

46. Numpy. Broadcasting. [Сайт] – URL: https://numpy.org/doc/stable/user/basics.broadcasting.html (Дата обращения – 06.05.2023)

47. Tirinox. Numpy – бродкастинг. [Сайт] – URL: https://tirinox.ru/numpy-broadcasting/ (Дата обращения – 06.05.2023)

48. Numpy. The N-dimensional array. [Сайт] – URL: https://numpy.org/doc/stable/reference/arrays.ndarray.html (Дата обращения – 06.05.2023)

49. Hexlet. Многомерные массивы в Numpy. [Сайт] – URL: https://ru.hexlet.io/courses/python-numpy/lessons/numpy-arrays/theory_unit (Дата обращения – 06.05.2023)

50. Документация Pandas. DataFrame. [Сайт] – URL: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html (Дата обращения – 06.05.2023)

51. Документация Pandas. DataFrame to numpy. [Сайт] – URL: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.to_numpy.html (Дата обращения – 06.05.2023)

52. Free Code Camp. Series and DataFrame in Python. [Сайт] — https://www.freecodecamp.org/news/series-and-dataframe-in-python-a800b098f68/ (Дата обращения — 11.05.2023)

53. Medium. Pandas data structures. [Сайт] — https://levelup.gitconnected.com/pandas-data-structures-7ac66072a8a (Дата обращения — 11.05.2023)

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты

Электронная библиотека по праву и экономике
https://право.информ2000.рф/uchebniki.htm

54. Pyviz. Python visualizations landscape overview. [Сайт] – URL: https://pyviz.org/overviews/index.html (Дата обращения – 06.05.2023)

55. Хабр. 50 оттенков matplotlib. [Сайт] — https://habr.com/ru/articles/468295/ (Дата обращения — 07.05.2023)

56. Хабр. Топ 6 библиотек Python для визуализации. [Сайт] — https://habr.com/ru/companies/otus/articles/558478/ (Дата обращения — 07.05.2023)

57. Medium. Deployment solutions for Python. [Сайт] — https://medium.com/anvil-works/streamlit-vs-gradio-vs-dash-vs-panel-vs-anvil-c2f86ad95ff3 (Дата обращения — 11.05.2023)

58. AWS. В чем разница между MySQL и PostgreSQL? [Сайт] – URL: https://aws.amazon.com/ru/compare/the-difference-between-mysql-vs-postgresql/ (Дата обращения – 05.05.2023)

59. ДАТА-ЦЕНТР Автоматика. [Сайт] – URL: https://dc.ru/ (Дата обращения – 04.05.2023)

60. ДАТА-ЦЕНТР Автоматика. О компании. [Сайт] – URL: https://dc.ru/about/ (Дата обращения – 04.05.2023)

Вернуться в каталог дипломов по менеджменту
https://право.информ2000.рф/management3/management3.htm
Полные тексты