**Министерство науки и высшего образования Российской Федерации**

**ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина»**

Кафедра «атомных станций и возобновляемых источников энергии»

Оценка работы\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Руководитель от УрФУ Велькин В.И.

Тема задания на практику

Прогнозирование энергопотребления в жилых зданиях с помощью машинного обучения.

ОТЧЕТ

Вид практики Производственная практика

Тип практики Научно-исследовательская работа

Руководитель практики от предприятия (организации)

**ФИО руководителя Подпись**

Студент Макамбаев Т.Х.

**ФИО студента**

Специальность (направление подготовки) 13.04.02Электроэнергетика и электротехника

Группа ЭНМ-281003

Екатеринбург 2020

## Принцип работы модели **прогнозирования**

Прогнозирование выполняется в несколько шагов.

1) Предварительный сбор и обработка данных

2) Прогнозирование выработки и потребления энергии

## Предварительный сбор и обработка данных

Данные являются краеугольным камнем для работы любой системы. Работа с данными требует большой ответственности, так как проблемы на данном этапе могут проявить себя не сразу, а на более поздних этапах. Поэтому, перед началом сбора данных требуется понимание, для чего они будут использоваться.

Во многих случаях, данные собираются в автоматическом режиме теми или иными датчиками, которыми располагает оборудование. Современные счетчики электроэнергии могут считаться умными, если они могут не просто измерять те или иные физические значения, а хранить и отправлять их в единую информационную структуру.

Фотоэлектрические панели, солнечные коллекторы, ветровые энергоустановки и прочее генерирующее оборудование подключается к контроллерам, которые позволяют следить за работой и состоянием оборудования. Продвинутые контроллеры автоматически собирают и сохраняют информацию для дальнейшего использования.

Помимо информации о состоянии оборудования необходимы данные, которые описывают состояние окружающей среды. Локальные метеостанции собирают информацию о температуре, влажности, давлении и пр. Такая информация имеет значения сама по себе, но в тандеме с информацией о текущей работе генерирующих установок она становится ключевой. Ключевой она становится, так как позволяет делает надежные прогнозы выработки энергии теми генераторами, которые напрямую зависят от состояния окружающей среды.

Данные нуждаются в обработке, так как они могут находится в разных форматах, собираться в разные интервалы и иметь разную точность, кроме того, в данных могут содержаться пропуски. Чем выше частота сбора данных, чем меньше пропусков и чем выше точность, тем точнее будет работать модель построенная на основе этих данных.

Зачастую, в собранных данных по той или иной причине могут отсутствовать данные за определенный промежуток времени, если отсутствуют данные за несколько минут или часов, то такой пробел не окажет существенного влияния. В некоторых случаях данные могут отсутствовать от нескольких дней, вплоть до нескольких месяцев.

Пропуски в данных можно исправлять различными способами, небольшие пропуски можно заменить средним или медианным значением из соседних измерений. Значения за несколько часов можно восстановить при помощи регрессионного анализа. В случаях особенно значительных пропусков, можно принять решение не использовать пропущенный промежуток данных вовсе.

Частота сбора данных различного оборудования может отличаться по тем или иным причинам. Высокая частота сбора позволяет создавать более точные и чувствительные модели. С другой стороны, при высокой частоте сбора, даже за небольшой промежуток времени накапливается значительный объем данных. Большой объем данных вынуждает приобретать больше накопителей, что приводит к большим экономическим расходам.

Помимо качества данных, их количество также имеет большое значение. В большом количестве низкокачественных данных мало смысла, в маленьком количестве качественных данных тоже.

В различных случаях, необходим разный период данных. Если случай затрагивает погодные условия, то необходимо иметь данные как минимум за один год. Год — это каждый месяц, такой промежуток позволит модели обучиться предсказывать погоду в любой месяц в любой день года. Использование лишь нескольких месяцев сильно ухудшит качество системы. Система будет успешно прогнозировать погоду в обработанные месяцы, а в остальные будет испытывать затруднения. Поэтому необходимо понять, какие могут быть зависимости в данных. Это необходимо для составления правильной картины ситуации и успешного построения системы.

Помимо обработки сырых данных, требуется подготовка данных для модели, так как разные модели требуют разные данные в зависимости от архитектуры модели. Кроме того, на финальный вид данных оказывает влияние компьютерные мощности, которые будут производить расчет. Чем мощнее компьютер, тем больше данных мы можем использовать за раз и тем точнее будут прогнозы модели.

Все вышеуказанные факторы оказывают влияние на сбор и обработку данных. В каждом случае требуется особый подход, который поможет получить требуемые результаты.

Для построения текущей модели требуются данные собранные с фотоэлектрических панелей ветровых установок. Помимо генерации требуются данные по потреблению электроэнергии и данные локальной метеостанции. Все данные взяты из открытых источников.

## Прогнозирование выработки и потребления энергии

Основными методами прогнозирования временных рядов или регрессий являются статистические модели. Благодаря высокой производительности компьютерных систем, наука машинного обучения, медленно развивающаяся с 1959 года, получила развитие в последние 10 лет. Нейронные сети позволяют получить результаты прогнозирования не хуже, чем при использовании статистических методов.

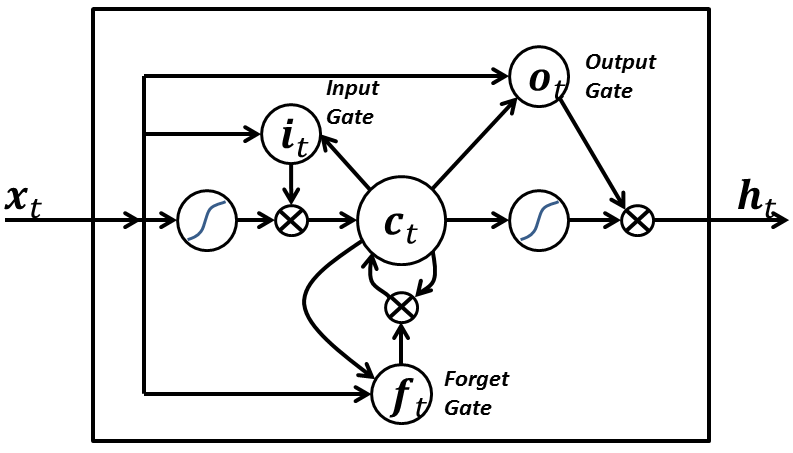
В текущем случае будет использоваться модель Long Short-term Memory (LSTM), разбор которой проводился в [18]. Данный тип рекуррентных нейронных сетей был предложен в 1997 году, но активное применение нашел лишь в последние несколько лет.

Модель выполнена в среде Python, с использованием библиотек Pandas, Numpy, Scikit-learn, Tensorflow, Keras.

LSTM является одной из разновидностью рекуррентных нейронных сетей. По сравнению с обычными нейронными сетями прямого распространения у рекуррентных имеются обратные связи. Благодаря этому, рекуррентные сети могут обрабатывать не только единичные данные, но и последовательные наборы данных. Поэтому рекуррентные сети часто находят применение при анализе непрерывных сигналов, куда могут входить: временные ряды, аналоговые сигналы, человеческая речь и пр.

Нейрон LSTM состоит из нескольких вентилей: входного, выходного и вентиля забывания. Схема нейрона приведена на рисунке 1.Такая структура нейрона позволяет бороться с проблемой затухающего градиента, которая возникает при обучении обычных рекуррентных сетей. Относительная нечувствительность к долгому отсутствию ранее увиденных сигналов является преимуществом перед обычными рекуррентными нейронными сетями.

Основными признаками, на основе которых строился прогноз выработки энергии для ФЭП, являются: температура окружающей среды, температура инвертора, температура модулей ФЭП, интенсивность излучения,

  
Рисунок 1 - Схема нейрона

относительная влажность, скорость и направление ветра, а также текущий месяц и час. Прогнозируемой величиной является количество энергии в Вт\*ч.

Для ветроэнергетических установок (ВЭУ) используются аналогичные параметры: температура окружающей среды, температура, давления воздуха, влажность, скорость ветра и направление ветра. В перечень временных признаков входят месяц, день и час. Прогнозируемой величиной является количество энергии в Вт\*ч.

Прогнозирование потребления электроэнергии в помещениях является аналогичной задачей, но в ней используются другие признаки, так как потребление зависит не только от состояния окружающей среды, но и других факторов, влияющих на поведение людей.