Smart Cooking temperature analysis system

Saharsh Shukla, Student & Mohit Bisht, Student

***Abstract*— With the advent of data science and the integration of it in every aspect of daily lives, it is easy to notice the stagnant home sciences that have not progressed since the late 90s. With the integration of Internet of Things into home appliances, it makes somehow easier to control these devices, but why stop at control? This paper explores the integration of simple sensor solutions to analyze data that can be used to further enhance and create functionality into connected devices with the help of Machine learning and cloud-based data collection, for the purpose of writing this paper, the cooking time of meat was explored.**

***Keywords—Internet of Things, Cloud, Data Science, Machine learning.***

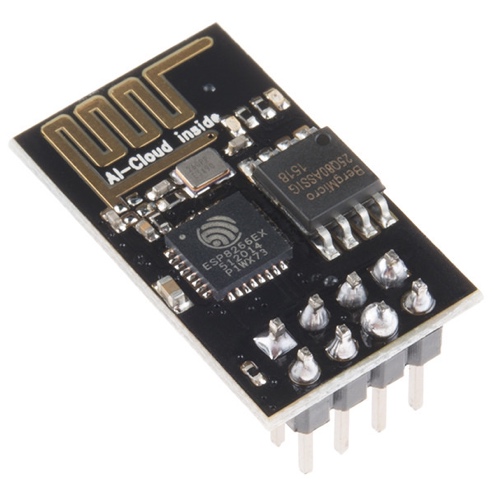
***Abbreviations—api – Application programming interface, SFrame – Size immutable Frame***

# I. INTRODUCTION

Most home appliances on the market in 2020 have some aspect of “Smartness” to it. Any smart appliance is either connected to the internet and provides control through smartphones or through web applications and/or provide predictions as to what a user might want. The second aspect of the so called “Smart” aspect of devices is still fairly stagnant in cooking appliances because of the open environment and varying climates. Most Ovens do not come with any extra sensors that help the user to benefit from the product.

The traditional approach to cooking relies on the intuition of the person using it. This project aims to aid the user by collecting accurate data and using analytics to get the perfect output or end product about 95% of the time with minimal variations.

For ease of understanding the paper is divided into sections that focus on different aspects of the project.



1Saharsh Shukla and Mohit Bisht are students at Vellore institute of technology in the SENSE department and are pursuing a Bachelor’s in Technology with Major in Electronics and communication Engineering

# II. METHODOLOGY

Diagram

Description automatically generated

## A. Hardware

|  |
| --- |
|  |

Currently the product is still under research, for the product to work perfectly it is important to note that a database of has to be developed for different articles to be cooked as well as the power, surface area and respective size of the closed environment that is to be used. The product shall be an array of smart sensors, namely temperature sensors, one as a probe and another as an ambient temperature sensor. The probe shall be used to measure the internal temperature of the article being cooked whilst the ambient temperature sensor shall be used to measure the external temperature or the environmental temperature of the oven itself, an IR sensor that will be responsible the distance and relative size of the article being cooked. The racks shall be fitted with Piezoelectric sensors to measure the weight of the article being cooked.

This data is relayed to a smart controller like an Arduino mega that can then send the data to cloud.

## B. Software

## 1) Initialisation

When the data has arrived, initially it shall be stored into the cloud, this data will then be read by a python file that will be using the Thingspeak api. The read api shall be used to download the data into the system which will then be loaded into an SFrame.

## 2) Data Cleaning

An Ipython notebook has been used in this experiment to determine if the concept is correct with very basic analytical tools. Once the data is collected into the system and initialized into an SFrame. This data is then plotted and checked and cleaned. For instance, Thingspeak shall provide the time of the data with a string format. We shall proceed to convert this data into DateTime format using the DateTime library and turicreate data visualization tool.

*3) Data anlysis*

Once the data is cleaned with the specific formats needed, duration can be calculated by subtracting the initial data time from the rest, this will be stored in a separate column called difference. This will be the duration at which the data was collected or time elapsed and will be very important in creation of the model itself.

# III. WORKING

This section will focus on the data that was collected and its visualization. (Refer to Tables section to see the format of the Data)

* The data used was collected through two sensors used with a Bosch HBF011BR0Z
* The following graphics are a part of the default visualization tools offered by ThingSpeak.

## A. Visualisation of Data

Graphical user interface, application

Description automatically generated

Figure 1 – External Temperatures – Thingspeak

Chart, line chart

Description automatically generated

Figure 2 – Internal Temperatures – Thingspeak

This data is in the format of CSV, it was then downloaded into the system using the Thingspeak API and a heat map was created signifying the most common temperature while cooking, this will show at what time the temperature is hardest to increase.

Chart

Description automatically generated

### Figure 3 – Heatmap – Internal Temperature vs External Temperatures

This showed us that the external temperature of 267(°F) was most common and internal temperature of 151(°F) was the most common. The TuriCreate Visualization Heat map tool was used for this. Once the heat map was created, further analysis showed us that the mean temperature was indeed infact 154(°F) for internal temperatures. The external temperature data analysis also suggested to confirm our output from the heatmap.

Using this initial data a model was trained by splitting the data into training and testing data.1

Further the loaded was cleaned to get the date and time correctly or to change its format from string to standard DateTime format.

Initial format - 2017-05-29 06:20:53 UTC

Output format - 2017-05-29 04:59:53+00:00

The output format could now be used for arithmetic operations using the DateTime library.

The initial time was recorded and the preceding times were then differed upon. This gave us the duration of operation.

The second model was trained with linear regression using the duration as a factor with internal and external temperatures.2

The Third model was trained with multiple regression using the duration as a factor with internal and external temperatures.3

**1,2,3 - Refer to Model part of the paper.**

## B. Visualisaiton of Model

This First, second and the third model were then compared with the original data to get the proof of concept.

## 1) Model 1 - Direct Linear regression with temperature as features

Chart, scatter chart

Description automatically generated Figure 4 – Model 1 plot for comparison.

The Orange dots signify the original data while the blue ones are the predictions of our models, We can see that the predictions are from accurate. This model is therefore not viable to use.

1. *Model 2 – Direct Linear regression with temperature and duration as features.*

Chart, scatter chart

Description automatically generated Figure 5 – Model 2 plot for comparison.

The blue dots signify the original data while the orange line is the linear regression. This shows that the model is better than the previous model but still lacks the accuracy that is desired from the final model.

1. *Model 3 – Multiple regression with temperature, duration and Difference as Features.*

Shape, scatter chart

Description automatically generated Figure 6 – Model 3 plot vs Model 2 plot for comparison.

The orange dots are the current model and its predictions while the green is the previous model and the blue dots are signify the original data.

The Visualization suggests that the this model is much more viable for use in commercial environments.

1. *Model 4 –Boosted trees regression for prediction of time*

Chart, scatter chart

Description automatically generated  
Figure 7 – Prediction of time or duration of run

The Orange dots signify the predicted run times whereas the blue dots signify the original data.

This model provides a much better outlook as to how duration can be predicted as is clear by the visualization.

## Model 5 – Autoregressive integrated moving average

Chart, histogram

Description automatically generated

Figure 8 – Prediction of time or duration of run

We find that the ARIMA time series analysis fails, this could be predicted before since ARIMA is essentially a linear regression type model.

ARIMA uses its own time lags to predict results.

A second difference was calculated to get a conversion of stationarity.

# IV. RESULT

We used three separate models for the prediction of temperature for the experiment. The aim was to get the proof of concept.

## A. Model 1

The reason the first model failed is an obvious one. Temperature is not an. Arbitrary value and increases slowly and steadily with time. When we do not include time elapsed into the model it fails at a very primary level. When we train this model, we expect to see the mean of the data predicted when supplied with a base data.

This holds true when we finally do get the output.

Table

Description automatically generated Table 1 – Table Model 1 with predictions.

As we can clearly see, the data is around 155(°F) as seen before in the heatmap. This is the Most common

Temperature in the operation.

## B. Model 2

Once we factor in the duration of the operation it is a

whole different story.

Table

Description automatically generated

Table 2 – Time Difference and duration

Since now the model can sense that there is a factor of linearity in increase of temperature it can predict much better. The difference in accuracy is also staggering, while the accuracy of the first model was a stagnant 12%, this model predicts with 67% accuracy which is a step in the right direction so to say.

Table

Description automatically generated

Table 3 – Model 2 Predictions

## C. Model 3

The previous model was still not good enough to be used in commercial application where the UX is the primary goal.

Linear Regression is therefore scrapped with much better Multiple regression models. This model not only predicts with a 95% accuracy, it also gives us proof of concept.

Table

Description automatically generated

Table 4 – Model 3 Predictions

The Maximum error has now decreased to the extent that this product can be used in commercial applications. The Maximum error was of 5% signifying more than 95% accuracy of prediction.

## D. Model 4 – ARIMA

We use Autoregressive Integrated Moving average for time forecasting. We try to check if a projection or a forecast for the data can be gained, this will be a game changer for the industry.

It could have been predicted that ARIMA will show a negative result since it is essentially linear regression, sure enough we get the following predictions.

A picture containing text, receipt, screenshot

Description automatically generated

# V. CONCLUSION

The data analysis and the third model provides proof that a product can be developed and can be used viably by manufacturers in their smart connected devices. A seamless integration with iOS and Android platforms using a coreML model with minimal delay.

The final product from the Point of View of the Consumer:

* The user cooks one article and stores it with a name.
* The data for this article is stored in the database and a model is trained and exported.
* The CoreML model is then used to predict temperatures and therefore the time taken to cook directly and gives a buzz alert when it is time to take out the article and stop the cooking process.
* This will therefore always lead to a perfectly cooked article.

## A. Data Visualisation

The data visualization was initially performed in a

developer environment using an iPython Notebook. In the final product the data analysis will be through base python files that upload and send the graphs to the servers, these graphs be retrieved and displayed in an interactive form to the user.

## B. Data anlytics

The data analysis can use better models that can increase the accuracy even further, for instance using neural networks instead of ML models might increase the accuracy of the model used. For now, a CoreML model was exported and will be coupled with a separate python file directly.

# VI. TABLE

Table

Description automatically generated Table 5 – Original Data

**An Arbitrary Date was used for the experiment**

## REFERENCES

1. https://blogs.mathworks.com/iot/2017/05/29/the-data-science-behindbbq/. *(references)*
2. https://apple.github.io/turicreate/docs/api/