**Predicting the Uncertainty of the Sales**

1 st Amit Singh

Department of Electrical Engineering

Stevens Institute of Technology Hoboken, NJ

[asingh61@stevens.edu](mailto:asingh61@stevens.edu)

2 nd Amit Singh

Department of Electrical Engineering

Stevens Institute of Technology Hoboken, NJ

[asingh61@stevens.edu](mailto:asingh61@stevens.edu)

Abstract—The goal of this study is to use hierarchical sales data from Walmart, the world’s largest company by revenue, to forecast daily sales for the next 28 days and to make uncertainty estimates for these forecasts. The data covers stores in three US States (California, Texas, and Wisconsin) and includes item level, department, product categories, and store details. In addition, it has explanatory variables such as price, promotions, day of the week, and special events. Several machine learning techniques were applied to problem, trained on a wide range of features, with varying success.

Index Terms—Artificial Neural Network, Linear Regression, Feature Extraction, Supervised Learning, Gaussian process, Uncertainty in Sales

1. **INTRODUCTION**

Predicting the sales of an organization play a critical role in the longevity of an organization. It may seem like a daunting task as predicting the weather, but both rely on the science and historical data. The sales data, store id, product code, location, time of the sales etc. holds a key to predict the future of sales. Understand the data and figure out the features which have more weightage and correlation in the sales. Is there a pattern in the sales? Once we have figured out a scientific way- machine learning way to predict the sales, what is the uncertainty of that prediction, estimate the levels of uncertainty. We have tried to look at uncertainty as a Bayesian process and modulate sales as an interval of uncertainty.

To accomplish, we have pulled the dataset from the Kaggle competition.

1. **TOOLS USED**

* Sklearn
* Keras
* TensorFlow
* Matplotlib
* Numpy
* Pandas
* Kaggle
* Statsmodels

1. **DATA COLLECTION**

The data for this project is collected from the Kaggle competition; <https://www.kaggle.com/c/m5-forecasting-uncertainty/data>

The data consisted of the following;

* Calendar.csv- This contains the calendar data and has the columns;
  + Date – the date
  + wm\_yr\_wk – week month of the year
  + weekday – Saturday or Sunday
  + wday – week of the day 1 to 7
  + month – month of the year
  + year – the year
  + d – an id for the date in d\_11 format
  + event\_name\_1 – weather the date is any event date
  + event\_type\_1 - weather the date is any event date
  + event\_name\_2 - weather the date is any event date
  + event\_type\_2 - weather the date is any event date
  + snap\_CA – snap data for CA
  + snap\_TX - snap data for TX
  + snap\_WI - snap data for WI
* Sales\_train\_validation.csv - Contains the historical daily unit sales data per product and store [d\_1 - d\_1913]
  + Id – Unique record for the row item
  + item\_id - Item id
  + dept\_id – department id
  + cat\_id – category id of the product
  + store\_id - store id
  + state\_id – State id
  + d\_1 - d\_1913 – day represented as d\_1 to d\_1913
* Sell\_price.csv - Contains information about the price of the products sold per store and date.
  + store\_id
  + item\_id
  + wm\_yr\_wk
  + sell\_price

1. Data preparation for the forcasting < DETAILS TO ADD>
2. Feature Engineering
3. Forecasting <Details to folloe>
4. Predicting Uncertainty-