



Effective Feature Engineering: A Structured Approach to Building Better Machine Learning Models

Dipanjan Sarkar



22 – 25 November, 2018



NIMHANS Convention Center, Bengaluru



Session Agenda



Introduction



**The Importance of
Feature Engineering**



**Case Study 1: NYC
Taxi Fare Prediction**

Hands-on Session



**Case Study 2:
Predicting
E-Commerce
Product Ratings**

Hands-on Session



Final Words



http://bit.ly/dhs18_ds

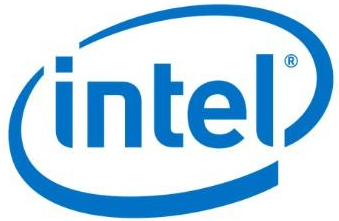
A blue wooden door with a diagonal brace and a central text overlay. The door is made of horizontal wooden planks and has a diagonal wooden brace running from the bottom left to the top right. The word "Introduction" is written in white, bold, sans-serif font on a dark blue rectangular background in the center of the door.

Introduction

About Me

Dipanjan Sarkar

Data Scientist, Author, Editor, Consultant & Mentor



Towards
Data Science



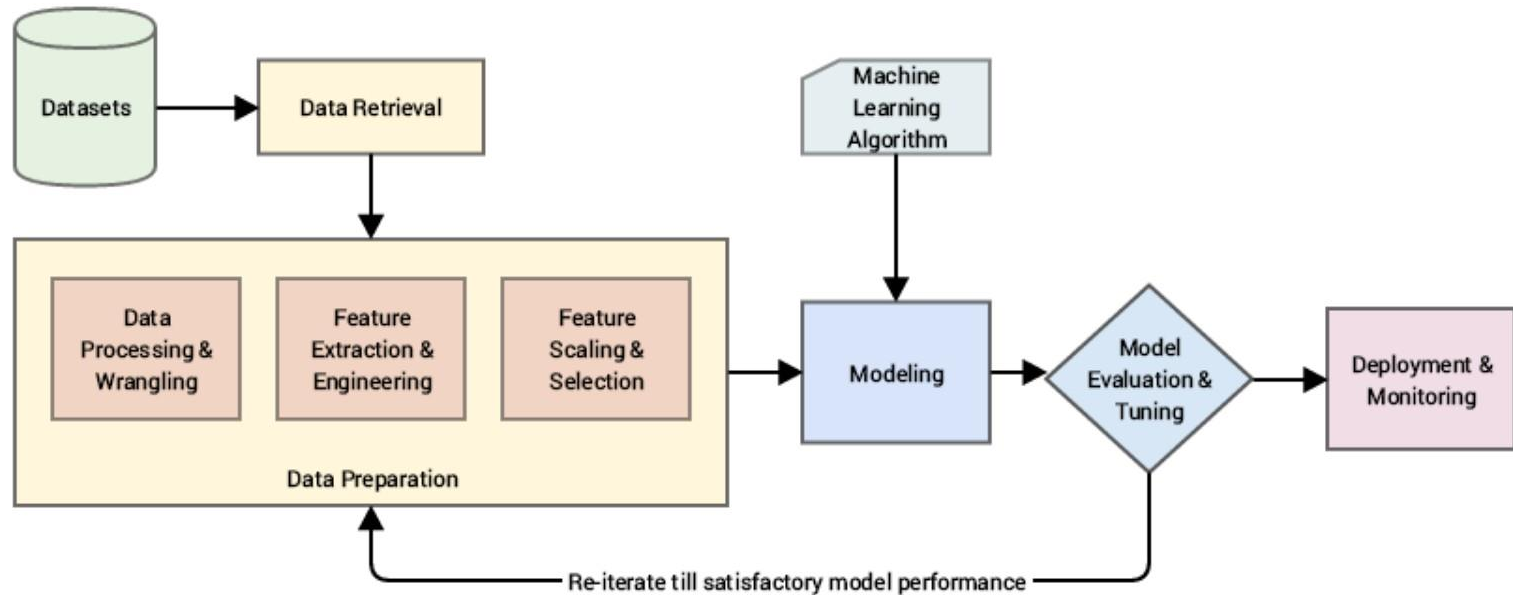
Intent of this Hack Session

- 1 Understand the Importance of Feature Engineering
- 2 Structured Workflow for building Machine Learning Models
- 3 Hands-on Case Study on a Regression Problem
- 4 Hands-on Case Study on a Classification + NLP Problem
- 5 Learn how to combine exploratory analysis with feature engineering

The background features a complex geometric design. Large triangular sections in dark blue, bright pink, and light beige are arranged to form a larger triangular shape. Overlaid on these is a pattern of rectangular blocks, similar to a brick wall, in shades of blue and pink. The text is centered within a dark blue rectangular area.

The Importance of Feature Engineering

Standard Machine Learning Pipeline



The Importance of Feature Engineering

- 1 Feature engineering, Data wrangling and Visualization are all aspects of Data Preparation
- 2 Creative and Innovative hand-crafted features can often be the deciding factor in winning competitions or deploying a model from a proof-of-concept to production
- 3 Feature Engineering is an art and requires domain knowledge and skill even with the advent of Auto-ML systems and Automated Feature Engineering tools
- 4 One of the most important stages in any machine learning pipeline

The Importance of Feature Engineering

1

“Coming up with features is difficult, time-consuming, requires expert knowledge. ‘Applied machine learning’ is basically feature engineering.”

— Prof. Andrew Ng.

2

“At the end of the day, some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used.”

— Prof. Pedro Domingos

3

“We spent most of our efforts in feature engineering. ... We were also very careful to discard features likely to expose us to the risk of over-fitting our model.”

— Xavier Conort

Case Study 1: New York City Taxi Fare Prediction



A photograph of a workspace, likely a desk, illuminated by a green desk lamp. The desk is cluttered with various items: two computer monitors (one displaying a colorful abstract image), a keyboard, a mouse, a soldering iron, pliers, a glass of water, and other electronic components. The background wall is green and has several papers or posters pinned to it. The overall atmosphere is that of a hands-on, creative, or technical session.

Hands-on Hack Session

A high-angle photograph of a clothing boutique. The floor is made of light-colored wooden planks. In the upper right, a white rectangular table holds several items: a folded blue garment, a piece of brown textured fabric, a small wooden box, and a collection of colorful, round, patterned buttons or pins. To the right of the table, a mannequin torso is partially visible, wearing a patterned top. In the lower left, a metal clothing rack is filled with numerous wooden hangers, each holding a dark-colored garment. The text "Case Study 2: Predicting E-Commerce Product Recommendation Ratings from Reviews" is overlaid in white, bold, sans-serif font on a semi-transparent grey rectangular background in the center of the image.

Case Study 2: Predicting E-Commerce Product Recommendation Ratings from Reviews



The image shows a laptop screen with a code editor. The left pane shows CSS code for a header, a main container, and navigation links. The right pane shows HTML code for a page layout with a header, a main section, and a footer. The status bar at the bottom of the editor indicates the current line and column.

Hands-on Hack Session



Final Words

Final Words

- Feature Engineering is the secret sauce to building great machine learning models
- Better features often beat building more complex models
- Always visualize the data as needed to understand the context
- Remember to always experiment and tune models on validation datasets
- Be aware and careful about data and target leakage
- Improvise hand-engineered features with automated techniques as needed



GETTING STARTED WITH NATURAL LANGUAGE PROCESSING

Interacting with artificial intelligent systems seems a bit simulated at times. This is because the way we converse as humans to one another is completely different from that we do usually with AI systems. Thankfully, research has been rampant in the area to bridge the gap in conversational AI systems. In this 8-hour workshop, you will get to know about natural language processing, creating word embeddings and developing learners to perform NLP tasks like sentiment analysis, auto correction and much more.



DIPANJAN (DJ) SARKAR

Data Scientist, Published Author &
Consultant



RAGHAV BALI

Data Scientist and Published Author

Contact Me

Interested in collaborating, research or writing?



LinkedIn

<https://www.linkedin.com/in/dipanzan>



GitHub

<https://github.com/dipanjans>

**Towards
Data Science**

Towards Data Science

<https://towardsdatascience.com/@dipanzan.sarkar>

Thank You

