

Effective Feature Engineering

A Structured Approach to Building Better Machine Learning Models

Dipanjan Sarkar

Data Scientist, Author, Editor Consultant & Mentor

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

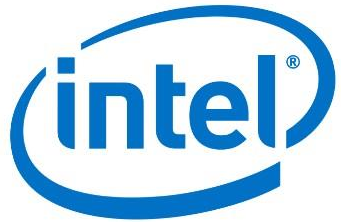
A blue wooden door with a diagonal brace and a central text overlay. The door is made of horizontal wooden planks and has a diagonal wooden brace running from the bottom left to the top right. The word "Introduction" is written in white, bold, sans-serif font on a dark blue rectangular background that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

Introduction

About Me

Dipanjan Sarkar

Data Scientist, Author, Editor, Consultant & Mentor



Towards
Data Science



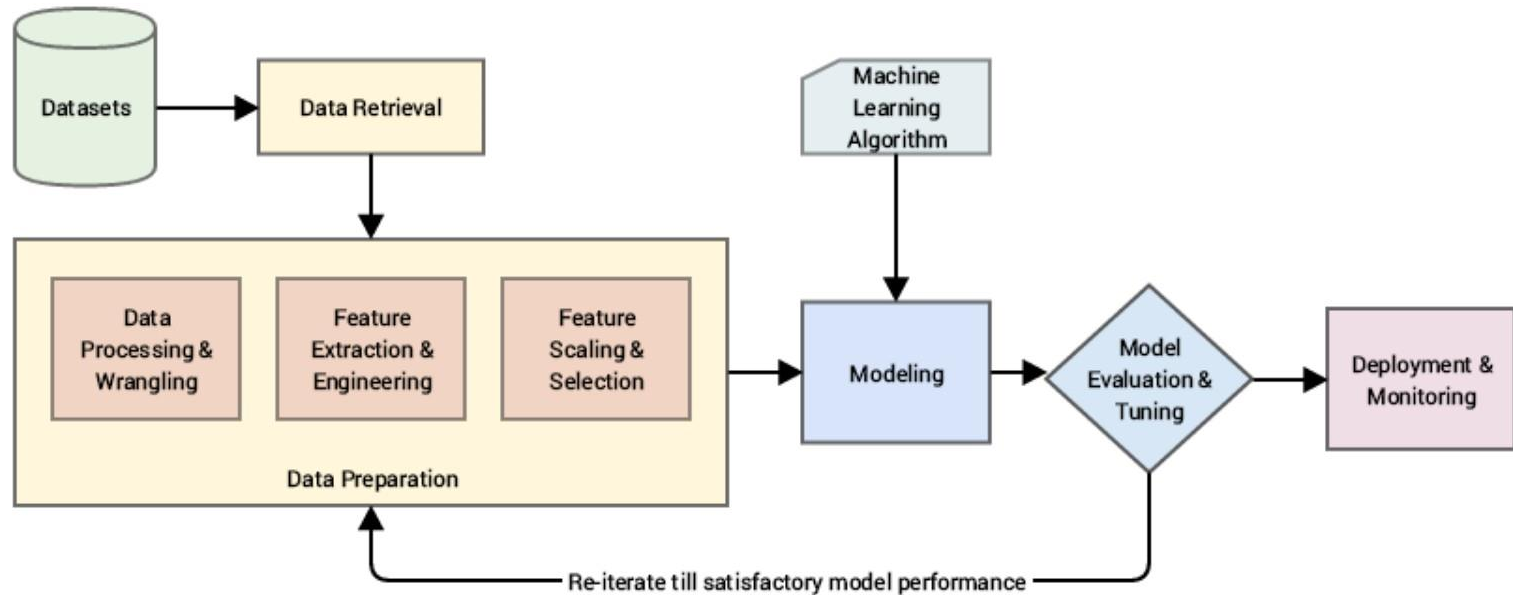
Intent of this Hack Session

- 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 and quadrilateral shapes in shades of blue, pink, and orange are layered over a background that resembles a brick wall. The text is centered within a dark, semi-transparent rectangular box.

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 high-angle, slightly tilted photograph of a yellow New York City taxi cab. The taxi is positioned in the lower right quadrant of the frame, moving towards the bottom right. It is a classic yellow cab with a red stripe and "TAXI" written on the side. The background consists of a dark asphalt road with white diagonal stripes forming a crosswalk. The text "Case Study 1: New York City Taxi Fare Prediction" is overlaid in a large, white, sans-serif font on a semi-transparent dark rectangular background that spans the middle of the image.

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, a glass of water, and several tools like pliers and screwdrivers. The background wall is green and covered with various papers and drawings. The overall atmosphere is one of a hands-on, creative, or technical session.

Hands-on Hack Session



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

A photograph of a laptop on a wooden desk. The laptop screen shows the Visual Studio Code editor with a dark theme. On the left, the Explorer sidebar shows a file tree with 'index.html' and 'index.css' selected. The main editor area is split into two panes. The left pane shows CSS code for a header and navigation links. The right pane shows HTML code for a page layout with a header, a main content area with a logo and list, and a footer with sign-up links. A status bar at the bottom of the editor shows 'Ln 23, Col 32, Spaces: 4, UTF-8, LF, HTML'. A semi-transparent dark grey banner with the text 'Hands-on Hack Session' in white is overlaid at the bottom of the image. In the background, a red portable power bank and a smartphone are visible on the desk.

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

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>