# 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**
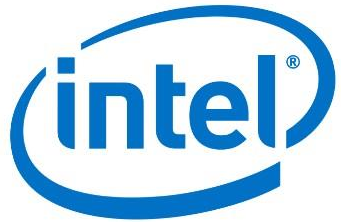
Hands-on Session

**Final Words**

# Introduction

About Me

# Dipanjan Sarkar

**Data Scientist, Author, Editor, Consultant & Mentor**

# 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 Importance of Feature Engineering

# Standard Machine Learning Pipeline
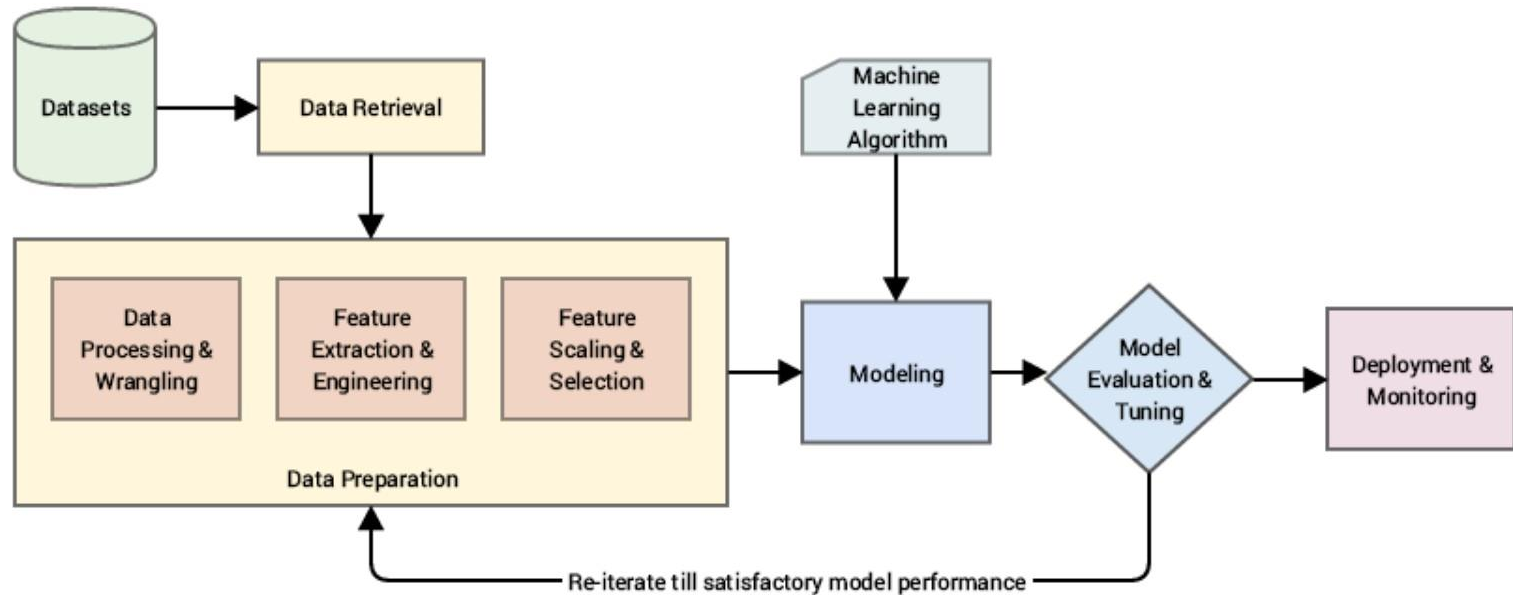
# The Importance of Feature Engineering

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

Hands-on Hack Session

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

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**

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