Course Project – Real-World Machine Learning Model

Machine Learning with Python: Zero to GBMs

Build a Machine Learning Project From Scr...

In the course project, you will apply the machine learning skills covered in this course by training an ML model on a real-world dataset. Follow these steps to complete your project:

1. Pick a large real-world dataset from Kaggle (see the "Resources" section below) and download it using opendatasets. Your training set should contain at least 50,000 rows and 5 columns of data.

2. Read the dataset description, understand the problem statement and describe the modeling objective clearly. You can also browse through existing notebooks created by others for inspiration.

3. Perform exploratory data analysis, gather insights about the data, perform feature engineering, create a training-validation split, and prepare the data for modeling.

4. Train & evaluate different machine learning models, tune hyperparameters and reduce overfitting to improve the model

5. Report the final performance of your best model(s), show sample predictions, and save model weights. Summarize your work, share links to references, and suggest ideas for future work.

6. Publish your Jupyter notebook to Jovian, make a submission below and share your project with the community. Optionally, you may also write a blog post and [contribute](#) to the [Jovian official blog](#).

There is no starter notebook for the course project. Please use the "New" button on Jovian to create a new notebook, "Run on Colab" to execute it, and `jovian.commit` to record versions. Please review the "Evaluation Criteria" and "Resources" sections below carefully before starting your project.

Submit   **Resources**   Discussions   Featured

Recommended Datasets

Use Kaggle to find [real-world datasets](#) and [past competitions](#) for machine learning.

Here are some project ideas to choose from:

1. Driver Alertness Detection - https://www.kaggle.com/c/stayalert

2. Insurance Claim Prediction - https://www.kaggle.com/c/ClaimPredictionChallenge

3. Financial Distress Prediction: https://www.kaggle.com/c/GiveMeSomeCredit

4. Used Car Quality Detection - https://www.kaggle.com/c/DontGetKicked

5. Photo Quality Prediction - https://www.kaggle.com/c/PhotoQualityPrediction

6. Bond Price Prediction - https://www.kaggle.com/c/benchmark-bond-trade-price-challenge

7. Biological Response Prediction - https://www.kaggle.com/c/bioresponse

8. Eye Movement Verification and Identification - https://www.kaggle.com/c/emvic

9. Next Song Suggestion - https://www.kaggle.com/c/msdchallenge

10. Xbox Game Suggestion - https://www.kaggle.com/c/acm-sf-chapter-hackathon-small

11. E-commerce Product Suggestion - https://www.kaggle.com/c/acm-sf-chapter-hackathon-big

12. Home Credit Default Risk Prediction - https://www.kaggle.com/c/home-credit-default-risk

13. Expedia Hotel Recommendation - https://www.kaggle.com/c/expedia-hotel-recommendations

14. Job Posting Recommendation - https://www.kaggle.com/c/job-recommendation

15. New Song Recommendation - https://www.kaggle.com/c/MusicHackathon

16. Non-Profit Donor Targeting - https://www.kaggle.com/c/Raising-Money-to-Fund-an-Organizational-Mission

18. Air Quality Prediction - https://www.kaggle.com/c/dsg-hackathon

19. Reducing Commercial Aviation Fatalities - https://www.kaggle.com/c/reducing-commercial-aviation-fatalities

20. Bulldozer Auction Price Prediction - https://www.kaggle.com/c/bluebook-for-bulldozers

21. Event Recommendation Engine - https://www.kaggle.com/c/event-recommendation-engine-challenge

22. Wind Power Generation Forecasting - https://www.kaggle.com/c/GEF2012-wind-forecasting

23. Bank Customer Category Classification - https://www.kaggle.com/c/elo-merchant-category-recommendation

24. Zillow Home Value Prediction - https://www.kaggle.com/c/zillow-prize-1

25. Santander Customer Satisfaction - https://www.kaggle.com/c/santander-customer-satisfaction

26. Insurance Decision Prediction - https://www.kaggle.com/c/prudential-life-insurance-assessment

27. Insurance Quote Conversion - https://www.kaggle.com/c/homesite-quote-conversion

28. Energy Load Forecasting - https://www.kaggle.com/c/global-energy-forecasting-competition-2012-load-forecasting

29. Product Launch Failure Prediction - https://www.kaggle.com/c/hack-reduce-dunnhumby-hackathon

30. Airbnb Booking Destination Prediction - https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings

31. Yelp Business Rating Prediction - https://www.kaggle.com/c/yelp-recsys-2013

32. Yelp Business Votes Prediction - https://www.kaggle.com/c/yelp-recruiting

33. Predicting Weather From Tweets - https://www.kaggle.com/c/crowdflower-weather-twitter

34. Inventory Demand Prediction - https://www.kaggle.com/c/grupo-bimbo-inventory-demand

35. Manufacturing Failure Prediction - https://www.kaggle.com/c/bosch-production-line-performance

36. Customer Business Value Prediction - https://www.kaggle.com/c/predicting-red-hat-business-value

37. PUBG Finish Placement Prediction - https://www.kaggle.com/c/pubg-finish-placement-prediction

38. Online Ad Demand Prediction - https://www.kaggle.com/c/avito-demand-prediction

39. Hotel Search Ranking - https://www.kaggle.com/c/expedia-personalized-sort

40. Webpage Evergreen Rating - https://www.kaggle.com/c/stumbleupon

41. User Detection from Smartphone Accelerometer Data -

42. Loyal Shoppers Detection - https://www.kaggle.com/c/acquire-valued-shoppers-challenge

43. Walmart Sales Forecasting - https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting

44. Grocery Sales Forecasting - https://www.kaggle.com/c/favorita-grocery-sales-forecasting

45. Insurance Policy Purchase Prediction - https://www.kaggle.com/c/allstate-purchase-prediction-challenge

46. Laptop Malfunctioning Parts Prediction - https://www.kaggle.com/c/pakdd-cup-2014

47. Mobile Ad Click-Through Prediction - https://www.kaggle.com/c/avazu-ctr-prediction

48. Auto Insurance Claim Prediction - https://www.kaggle.com/c/porto-seguro-safe-driver-prediction

49. Taxi Trip Duration Prediction - https://www.kaggle.com/c/nyc-taxi-trip-duration

50. African Soil Property Prediction - https://www.kaggle.com/c/afsis-soil-properties

51. Search Result Relevance Prediction - https://www.kaggle.com/c/crowdflower-search-relevance

52. West Nile Virus Detection - https://www.kaggle.com/c/predict-west-nile-virus

53. ECommerce Product Classification - https://www.kaggle.com/c/otto-group-product-classification-challenge

54. Windows Malware Detection - https://www.kaggle.com/c/malware-classification

55. Hourly Rain Prediction - https://www.kaggle.com/c/how-much-did-it-rain

56. Shopping Trip Type Classification - https://www.kaggle.com/c/walmart-recruiting-trip-type-classification

57. Context Ad Click Prediction - https://www.kaggle.com/c/avito-context-ad-clicks

58. Duplicate Ads Detection - https://www.kaggle.com/c/avito-duplicate-ads-detection

59. Instacart Market Basket Prediction - https://www.kaggle.com/c/instacart-market-basket-analysis

60. Russian Real Estate Price Prediction - https://www.kaggle.com/c/sberbank-russian-housing-market

61. Santander Customer Transaction Prediction - https://www.kaggle.com/c/santander-customer-transaction-prediction

NOTE: It is not compulsory to use one of the above datasets. You can select a dataset from any online source of your choice.