DataGenie

Discover Autonomous Data Intelligence

Data Science Hackathon 2025

Overview

DataGenie started in 2020 with the intention of doing things the smarter way - to automate the boring and letting innovation be rightfully at the center.

We are a geographically distributed team passionate about revamping the way analytics is consumed in the enterprise today. Behind the DataGenie magic is a hand-picked team of experts that has over 80 years of combined experience in data analytics. We have first-hand experience with both the ‘good’ & the ‘not so good’ parts of the current analytics landscape. While we retain & enhance the ‘good’, we know how exactly the ‘not so good’ is to be revamped.

At DataGenie, our vision is to make it effortless for businesses to realize data-driven decision-making at scale. We are agile, highly collaborative, and customer-centric in our approach to solution engineering and technology.

There's a lot that goes into building a culture and we keep it simple - work on what drives you and solve real problems alongside some of the smartest people in the business. It's real every step of the way.

We invite you to participate in our Hackathon themed ***“Data Alchemy - Transforming Raw Data into Golden Insights”***.

Put on your creative hats and get Hacking!

Problem Statement

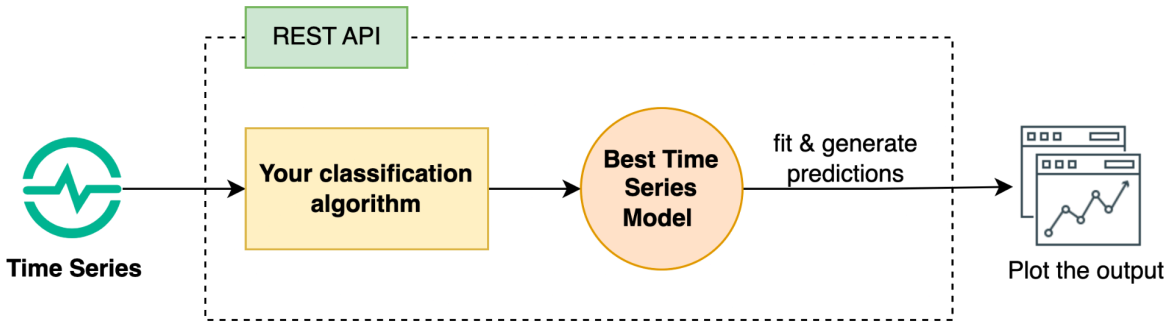
We're in a digital economy where data is more valuable than ever. It's the key to the smooth functionality of everything from the government to local companies. Without it, progress would halt.

Given the importance of data and analytics in today's digital economy, it's essential to develop intelligent analytics systems that can help organizations make sense of the massive amounts of data they collect.

At DataGenie we do this at scale and provide automated easy to consume insights to business users. To achieve this we have many modules working together and one of the components is the anomaly detection module where we use time series analysis models to forecast and detect anomalies.

To trigger curiosity and the critical thinker within you, we at DataGenie have decided to formulate this insight detection problem into a hackathon.

The goal is to create an efficient time series model selection algorithm



The above diagram explains the details on what you should be implementing:

1. **Checkpoint 1: Classification Algorithm** - train a classifier model which is capable of predicting which is the best time series model to use from the model suite based on the given time series data. **The best model is the one which results in the lowest MAPE value.** *Hint: Extract time series features*
2. **Checkpoint 2: Generate predictions** - use the **pre trained classifier** to obtain the best model. Split the given time series into train and test. Fit the model on train data and **generate predictions for test data** and also calculate the **MAPE** value.

3. **Checkpoint 3: REST API** - a simple REST API which allows the user to upload any time series data and get predictions for it along with the details of which time series model was used and MAPE value.
   Note: You can use ***sklearn.metrics.mean_absolute_percentage_error*** to calculate the MAPE value.
4. ***(Bonus)*** **Checkpoint 4: Simple UI & Deployment** - deploy a basic UI which allows users to upload a time series data (in a specific format) and obtain a plot. The UI should use the REST API you created underneath. *Hint: Use plotly for plotting*

Notes

- A clear explanation of how the classifier is trained has to be illustrated either as part of the README or as a video.
- If required you can use Deep Learning to develop the classifier
- The UI just needs to be functional and **will not** be considered for evaluation
- Make necessary assumptions where ever required and mention them in the README
- You are free to use any time series data for training the classifier. **You can find few sample time series data here: Sample Time Series**

Sample REST API request & response

POST */predict?date_from=2021-08-01&date_to=2021-08-03&period=0*

- date_from - date from which you will generate predictions
- date_to - date until which you will generate predictions

```
[
    {
        "timestamp": "2024-01-01T00:00:00",
        "point_value": 123
    },
    {
        "timestamp": "2024-01-01T00:00:00",
        "point_value": 131
    },
    {
        "timestamp": "2024-01-01T00:00:00",
        "point_value": 293
    }
]
```

```
// response
{
    "forecastability_score": 8.4,
    "number_of_batch_fits": 23,
    "mape": 0.2,
    "avg_time_taken_per_fit_in_seconds": 1.2,
    "results": [
        {
            "timestamp": "2024-01-01T00:00:00",
            "point_value": 123,
            "predicted": 123.5,
            "is_anomaly": "no"
        },
        {
            "timestamp": "2024-01-01T00:00:00",
            "point_value": 131,
            "predicted": 134,
            "is_anomaly": "no"
        },
        {
            "timestamp": "2024-01-01T00:00:00",
            "point_value": 293,
            "predicted": 128,
            "is_anomaly": "yes"
        }
    ]
}
```

Rules

1. This is an **individual participation** hackathon
2. It's an **online hackathon**, you can participate from anywhere.
3. Upon receiving the problem statement you will have **4 days** time to submit your implementation.
4. Use **Python** as the primary programming language for the implementation.
5. Any API written should follow REST Principles.
6. Use **FastAPI** (https://fastapi.tiangolo.com/) as the backend framework for the APIs.
7. The hack must be made available on **GitHub as a private repository** giving DataGenie access to read. Provide read access to **github usernames: (KrishnanSG & Starkyv)**
8. To be eligible for evaluation you must have completed **at least 3 checkpoints**.
9. The repository **must contain a README.md and short video** explaining the implementation; it could contain diagrams or could have instructions on how to set up and use it.
10. **Any plagiarism detected will result in immediate disqualification.**
11. **Your hack must be developed entirely during the Hackathon duration.**

Evaluation Criteria

1. General coding standards & usage of git
2. Features used for training the classification model
3. Classifier implementation and explanation of how the classifier was trained
4. Data science & time series analysis knowledge
5. Working & Demo of the application

Happy Hacking!
Cheers from DataGenie Team 😀

FAQs

1. **Do I need to pay any money to register for the Hackathon?** No. You do not have to pay anything to anyone to register yourself.
2. **Do I need to have any specific qualifications to be a participant for the Hackathon?** If you love data science and know how to code, you are more than welcome to participate in the Hackathon.
3. **How do I submit what I have made for the Hackathon?** You have to develop the application on your local system and commit the source code to a private GitHUb repository giving DataGenie access along with instructions to run the application.
4. **How is the environment? Will your environment support any language? Will you provide any IDE and DB for us to work on ideas?** You have to develop the entire software application on your local system.
5. **Does one have to be online and available for the entire duration of the Hackathon?** No, you can develop the application on your local system based on the given problem statement and then submit it.
6. **Do I need to give a demo for the product that I have built?** Yes, a short video explaining your approach and your implementation.
7. **Is there a constraint on the frameworks I can use?** For the frontend you are free to use any framework of your choice but for backend you are expected to use **Python** programming language and **FastAPI** (https://fastapi.tiangolo.com/) as the backend framework for the APIs.
8. **What ML or DL Frameworks can I use?** You are free to choose the frameworks of your choice.
9. **I have a question that is not listed here. Who can I contact?**
   You can send an email to ***hackathon@datagenie.ai*** with your queries.