

PROBLEM STATEMENTS FOR ERR_404 4.0

Problem Statement Code: 001

compliance-automated.io

Company Name: compliance-automated.io

Domain: Machine Learning

Introduction:

Security controls are documented in Compliance documents such as policy, SOP's, security plan and address specific compliance framework such as PCI DSS, SOC etc. These compliance documents are long, repetitive, and tedious to read and understand. Further challenges are on the Auditors to wrap up testing of security controls in an extremely short amount of time without missing anything significant such as a critical security tool, process, or role responsible for defining, deploying, or maintaining compliance for each specific security control. Some compliance frameworks and the resulting word documents that demonstrate compliance posture of a cloud hosted application (SaaS, PaaS, IaaS) can be extremely complex - spanning 17 unique control families, 650 + pages. That's the first part, 2nd part is to independently examine that what was documented has been "actually" deployed in production, for e.g., a SaaS claiming that it has 4 distinct roles for its end user and has 3 Admins Roles in AWS IAM or 2 AD Accounts for SaaS backend Admins. We would like this to be as much as automated.

Expected Solution:

Write a ML algorithm to parse the raw word document such as security plan or SOP, then depending upon the title of security control generate training data (we might provide them), develop a ML model to extract key features from the raw word document. Finally, create raw API calls to AWS IAM or Powershell scripts to query AD or SaaS Application to verify the "facts" extracted by the ML model. We will try to provide the standard set of APIs to execute this task.

Company Name: compliance-automated.io

Domain: Application Development and Security

Introduction:

Numerous customer service scenarios can occur through multiple channels such as web applications (online self-servicing) or mobile apps or chat bots or face to face meeting or through phone calls. Regardless of the channel, typically these interactive sessions depend on core enterprise applications and involve data input from one or more parties (end user, customer service representative or CSR & other 3rd party as needed for a business use case). This data input is a critical step of any business process, mistakes at this step can be extremely costly for all the parties specially corporations. Often the end customer does not see the application that a CSR is using on the other end of the interactive session. What if the end user can see what the CSR has done (regardless of channel used) after the information needed has been captured?

Expected Solution:

Given few static web pages (HTML, CSS etc.) from an enterprise application (web app, mobile app, phone calls/IVR, chat bots etc.) - Develop QR code or single page web app, that will accomplish the following:

- The end user will see a limited number of fields with populated data fields.
- The end user can correct any mistakes.
- The end user will approve the steps taken so far by CSR.
- The end user will hand the process back to CSR by closing the QR based single page web app.
- Solution should provide persistent storage of these interactions by saving this in a no SQL database so in future the end user or CSR can precisely trace back the process step or query it

Problem Statement Code: 003

Company Name: Jubi.ai

Domain: Machine Learning



Introduction:

Given one input, a general search term such as Mutual Fund, Union Budget 2021, LIC, UPSC, Elon Musk etc. The aim is to find all the questions being asked by the public online in recent days related to the searched term. The data scraped is to be cleaned, classified, and sorted according to relevance and frequency of the question and stored in suitable format or displayed on the browser. Output could be JSON, Excel or List, whichever is feasible.

Expected Solution:

You can use any source to accumulate data like Google, Quora, Twitter API, which allow you to scrape data using selenium in python or any other technology. The system could be an API or a web-based application but not a mobile application, something which can be easily integrated in any existing system. The main aim is to be able to sort questions based on relevance and frequency; The question which is most relevant to the input, and has been asked the most, should be at the top. Also cluster related data using a machine learning model which can apply a certain heuristic to classify similar questions. For example: "What is Elon Musk's date of birth" and "When was Elon Musk born" should be classified under the same category or be treated similar.

Problem Statement Code: 004

Company Name: N/A

Domain: App Development & Machine Learning.

Introduction:

The World Health Organisation (WHO) has declared the coronavirus disease 2019 (COVID-19) a pandemic. A global coordinated effort is needed to stop the further spread of the virus. The best way to stop the spread of covid 19 is to have the citizens themselves be aware of the areas that have a huge amount of covid-19 cases / patients or are containment zones and not go to those areas unless absolutely necessary. It is difficult for a citizen to keep track of such information and hence a solution is needed to do the same.

Expected Solution:

Develop an application that will accomplish the following goals:

- Visualize the areas affected by covid-19 on a map
- Show various statistics / data for spread of covid-19.
- If the user enters or is in proximity to an area that is a danger zone they should be notified of it.
- Predict the next risky zones and / or how covid-19 will further spread using past data.

Problem Statement Code : 005

Company Name: Bitkraft

Domain: Web Development



Introduction:

Process the sample data to generate GST invoices for customers by creating the below 2 schemas. Each invoice constitutes the following.

Important Information:

1. Every invoice has multiple lines in the data sample and is identified by unique values in the ref_id column
2. All Taxable lines are marked as "Taxable" in the taxable_status column
3. All Non-Taxable lines are marked as "Non-Taxable" in the taxable_status column
4. All Taxes lines are marked as "Taxes" in the taxable_status column
5. Every line is a part of a group and the group is identified by the charge_code_component column.

Header Information:

1. Invoice No [ref_no]
2. Customer name [First Customer Name of an invoice]
3. Departure Airport [from_airport]
4. Destination Airport [to_airport]
5. Invoice date [charge_date]
6. Total Value of the invoice [booking_charge_amount]
7. Total Taxable value [booking_charge_amount and taxable_status]
8. Total Taxes value [booking_charge_amount and taxable_status]
9. Total Non-Taxable Value [booking_charge_amount and taxable_status]

Line Level Information:

All invoices needed would have multiple lines based on the number of **unique charge_code_components** values. All such lines need to have the following

1. Charge_code_component Name
2. Charge_code_component Taxable or Non-Taxable Value
3. Charge_code_component Taxes Value

Expected Solution:

Use the mentioned above 2 schemas and data in them, after adding the data , create a page to search for an invoice by invoice number and generate a PDF for the invoice which includes header information and line information as per the sample attached.

Sample Data and Output of the Schema is provided.

Problem Statement Code: 006

Company: Kasmasoftware

Domain: Web Development



Introduction:

“The Airline Group” or “TAG” is a consortium owning a group of airlines in Europe. The company has around 50,000 employees across the airline is up. TAG has its own IT division with its branding – “TAG tech”. “TAG tech” handles all the IT operations across the airline group.

TAG tech has 3 major functions:

- Managing Business-As-Usual (BAU) operations of the various applications running on its infrastructure
- Track and fix issues and bugs
- Programs / Projects to align with the companies strategy/vision

Managing Business – As – Usual (BAU) operations

The functioning of business applications that support day to day operations is critical to the very existence of the organization. They are also called business-critical applications. The IT infrastructure (servers, networks, support, etc) are the backbone of these applications. This function tools/agents the existing IT infrastructure can support the optimal running of these applications. This function can include application end users, helpdesk support teams, application maintenance teams, monitoring tools/agents.

Track and fix issues/bugs

IT systems (hardware/software) need to have efficient issue tracking and fixing mechanisms to ensure business continuity. This function continually monitors bugs/issues reported by the users or system monitoring tools and tracks them to completion/issue resolution. The organization has employed ample resources to ensure any issues in its business-critical systems are promptly investigated and fixed.

Deliver Programs / Projects to align with the companies strategy/vision

Every organization needs to keep up with the changing market conditions and rapid technological advancement. To cater to this, TAG tech has this major function of initiating, managing, and delivering software projects that would keep the organization aligned to its mission statement as well as keep up with the market trends and latest technologies. TAG tech implements the PRINCE2 & Agile project management methodology to manage its portfolio.

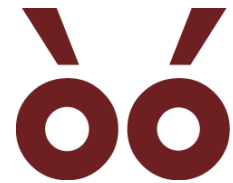
The **ITSM (Information Technology Service Management) system** ensures that it caters to all the needs that are required by “TAG - Tech”.

Expected Solution:

You are required to develop a web based application that performs the following:

- Whenever a FDR is created by the business user it is displayed in the dashboard for the proposal lead and the project manager.
- The business user should only be able to see the FDR that he/she has created.
- The FDR should have the following fields:
 - FDR Type (Solution Sketch or Statement of Work)
 - The operating company (You can use dummy values here, at least 4 companies are needed)
 - Project Number
 - Department of the project
 - Project Title and Description
 - Name of Proposal Lead
 - Status of the FDR (Open or Closed)
 - Costing Sheet (the format is available [here](#))
 - A field to upload the Solution Sketch
- The business user can only create the FDR and should view the status of the FDRs.
- The proposal lead can update the FDR
- The project manager can only view the FDR and add the SMEs
- The SMEs can update the information regarding the FDR.
- You need trackers to track the progress of the FDR, namely
 - Solution Sketch Tracker (Visible to Proposal Lead)
 - Triage Tracker (Visible to IT Infrastructure, Proposal Lead, Project Manager)
 - Statement of work action Tracker (Visible to Proposal Lead)
- The workflow of the entire Application can be viewed [here](#)

Problem Statement Code: 007



Company: Snoop Technologies

Domain: Machine Learning, Web Scraping & NLP

Statement:

Design & Develop an API with Web Interface to Extract/Create a good Call To Action Statement. The system should take few inputs to define the clarity of the type of CTA Statement you desire and on processing the inputs the system should produce a list of good to use CTA Statements.

Explanation/Examples:

Let's take for example you need a CTA statement for food ordering. Now let's consider the system needs two inputs: Type of CTA & Query Keywords respectively. Just for example, consider there are types of CTA like Order, Sell, Open, Browse, Sign Up, Subscribe, Engage, Remind, etc. Now for this particular use case you pick Order as your type of CTA and you enter Query Keywords: Food, Hunger. Then the expected output should look like the following:

- Feeling Hungry? Tap to Order Now!
- Hungry? Tasty food just a click away!
- Why Control your Hunger Pangs, Order Now!
- Hungry but Health a priority? Clean & Tasty Food Just a click away!
- Etc..

Some More various examples of statements:

- Bet you can't shop just one! Stress free styles @ min 50% off [CTA: Sell, keywords: shop, stress free]
- Your Determination to procrastination is commendable! [CTA: Remind, keywords: procrastination]
- On a weekend Binge Watch? Click to Subscribe Now! [CTA: Subscribe, keywords: Binge, Movie]

Proposed Flow(Not Compulsory for Hackers): Take/Make a dataset of good CTA's with its respective type of CTA & Query Keywords Respectively and feed it onto a Machine Learning Training module for training purpose. On receiving a query you must scrape appropriate sentences from the web. And finally use NLP & Machine Learning for filtering out only the relevant Statements and display the output. Voila!

Problem Statement Code: 008

Company: Koshex



Domain: Machine Learning and Web Development

Introduction:

The idea is to help people view and study their overall spend analysis by developing a simple web app to analyze all the purchases made by scanning the credit card and bank statements.

Expected Solution:

Use any method to parse the PDF and track payments. Using AI and machine learning, the app should be able to group items category-wise, for example, food, clothes, fuel, etc. on a date range view i.e weekly/monthly/yearly The solution should run on at least 5 different bank and credit card statements e.g. ICICI, HDFC, SBI, YES, AXIS

Additional Requirements: (If Time permits) The app should prompt users when they overspend or make repeat purchases

Note: If you need bank and credit card statements, please let us know. We'll arrange it.