DATA AND ANALYTICS DAILY DIGEST

Your daily feed, curated from all over the web!

https://www.linkedin.com/in/suva/

How to Implement a Data Privacy and Protection Strategy for Remote Teams

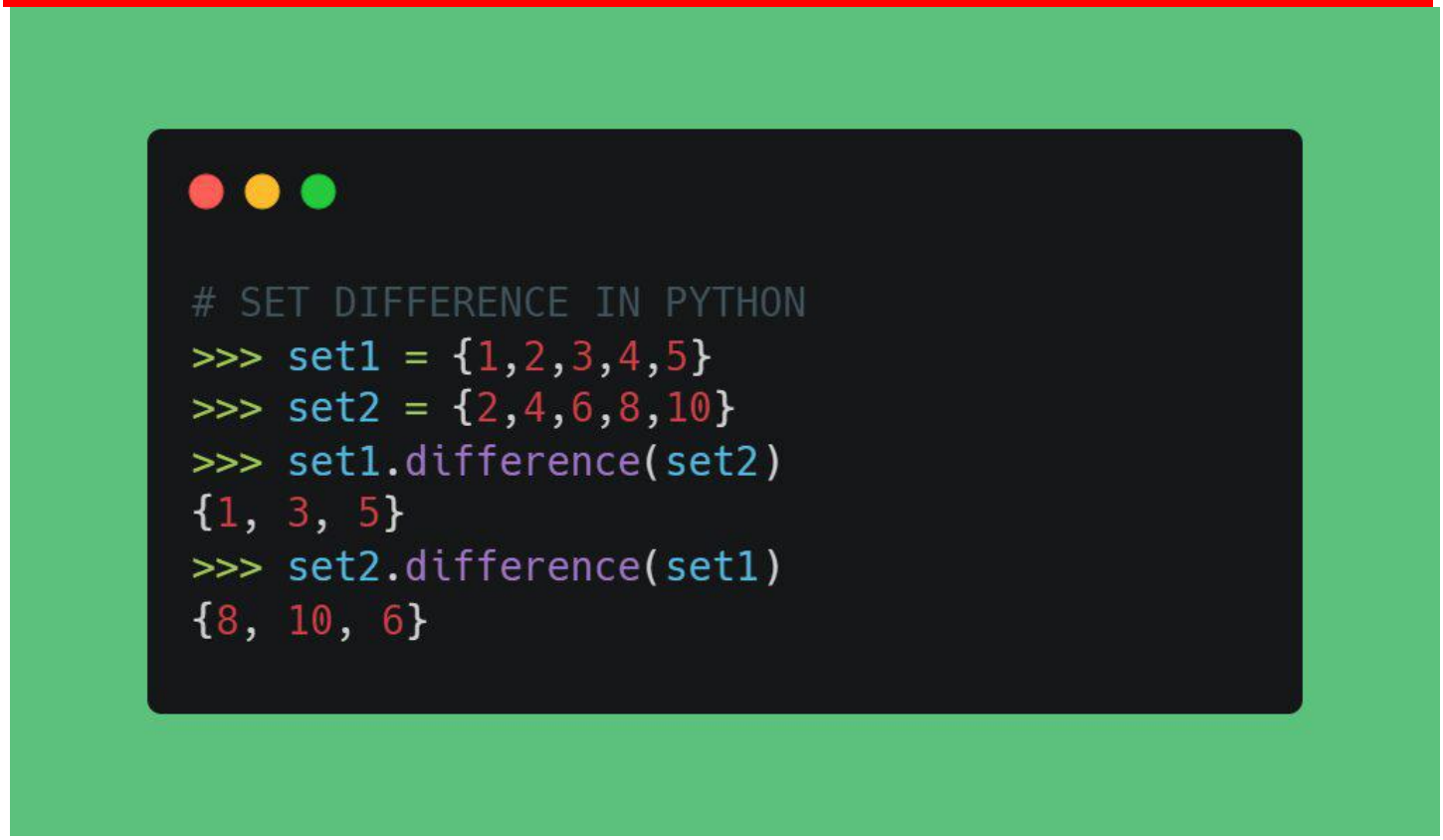
In this comprehensive guide, well walk through a simple five-step process for implementing a robust data privacy and protection strategy for your remote team. Once you have a strong grasp of all your current risks, its time to develop a comprehensive data privacy and protection policy for your organization. But a robust data privacy and protection policy is necessary to help secure sensitive information and create a foundation for effective metrics. Start implementing a data privacy and protection strategy for your remote team today and safeguard your companys future.

Read full article on www.datasciencecentral.com

5 More Command Line Tools for Data Science

Use these tools to Access API, Manipulate CSV files, download datasets, and more from your terminal. After installing, just type ipython in the terminal and start performing data analysis just like you do in Jupyter notebooks. In the example, we are downloading the machine learning data from the University of California and saving it as a CSV file. In this article, we have learned about CLI tools to download the dataset, manipulate it, perform analysis, run scripts, and generate reports.

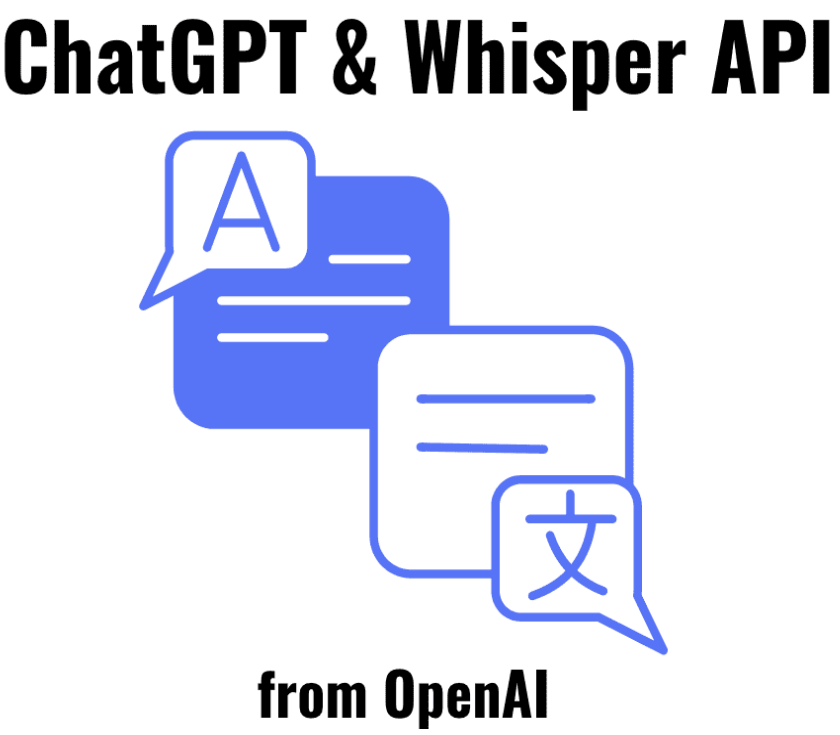
Read full article on www.kdnuggets.com



How to Find Set Difference in Python

Learn how to find the set difference between two Python sets using the difference() method and the difference (-) operator. You will learn how to use the built-in set method difference() and - operator to find the set difference, and how to debug common errors during the process. Before computing set difference on Python sets, let's quickly review the set difference operation. To find the set difference, we can use either the difference() method on a Python set or the - operator.

Read full article on www.kdnuggets.com



New ChatGPT and Whisper APIs from OpenAI

OpenAI has made its ChatGPT and Whisper models available on its API, allowing developers to have access to AI-powered language and speech-to-text capabilities. Developers will now be able to use OpenAIs open-source Whisper large-v2 model, which provides much faster and cost-effective results. Rather than having to use OpenAIs current language approach, ChatGPT and Whisper APIs will allow third-party developers to easily integrate them into their platforms. The launch of ChatGPT and Whisper APIs is expected to have a profound impact on the community of developers.

Read full article on www.kdnuggets.com

Significance of IVR in VoIP Communications

With the help of the custom IVR software development services provided by VoIP companies, any business can obtain it. Redirects with minimal hold time: Calls are routed to the appropriate department or agent thanks to an IVR system, which lets callers connect with the relevant person. IVR, on the other hand, functions as a self-help tool to provide customers with solutions without them even speaking to a live agent. Your phone number and an IVR menu can be used by callers to get basic information outside of usual business hours-or while agents are occupied with other calls.

Read full article on datafloq.com

Worlds leading Accounting Firms 2023

A professional who wants to excel in his/her career in accounting and finance always looks for top firms that offer innumerable and elated job opportunities with attractive and high paychecks. It is also a part of the Big 4 Accounting Firms, which are Deloitte, PwC, EY & KPMG. Ernst & Young Global Ltd., popularly known as E&Y or just simply EY, is one of the biggest as well as third largest professional services firms in the world. Binder Dijker Otte, also called BDO, is another biggest accounting and other professional services organizations.

Read full article on datafloq.com

Why Does KYC Matter For Fintech Companies?

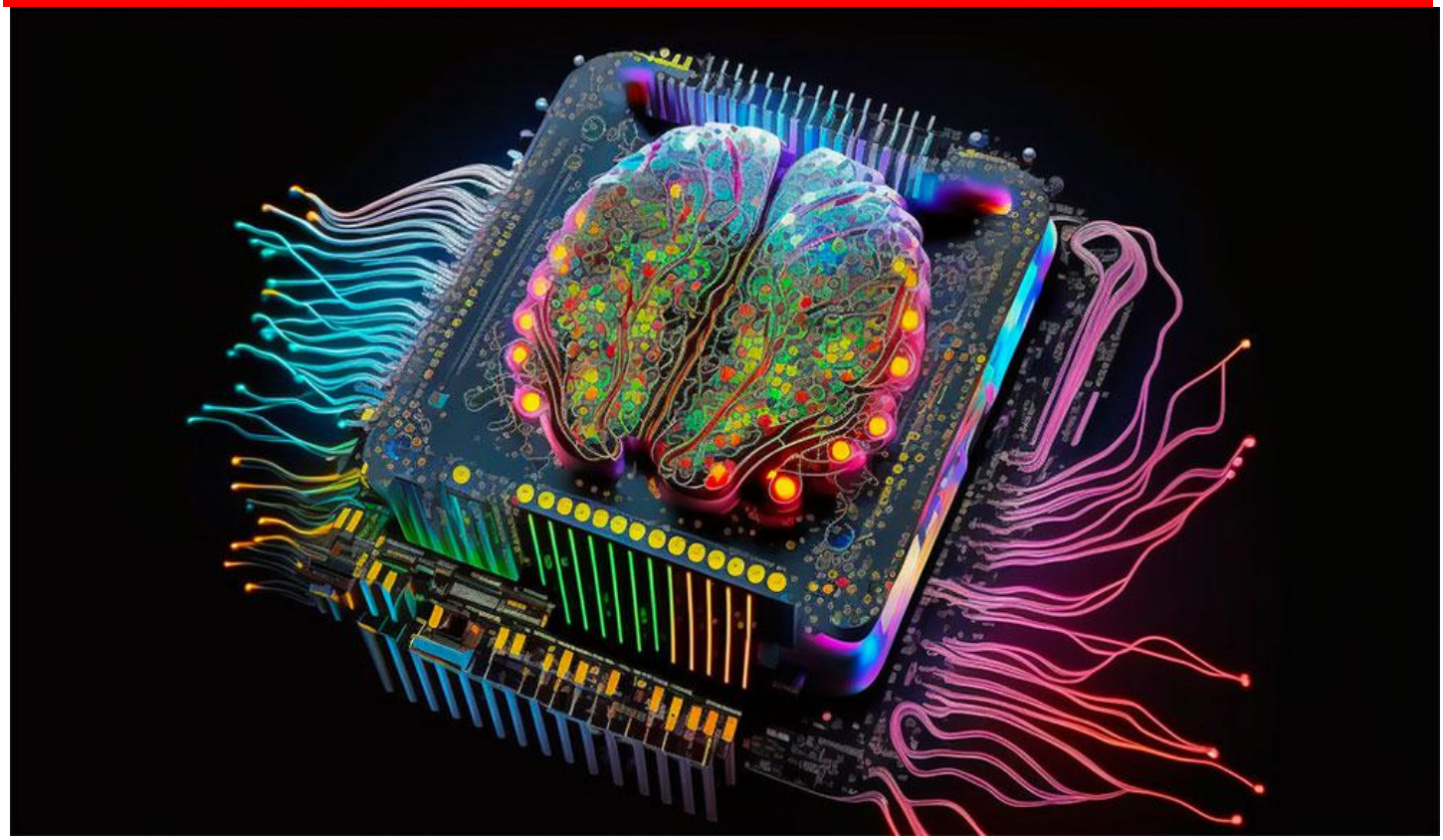
When it comes to fintech companies, knowing who their customers are, verifying that they are who they claim to be and making sure that their platform isnt misused to commit fraud of any kind is vital to maintaining a brand image and complying with government regulations. Irrespective of the jurisdiction and local regulations, a KYC program can be broken down into 3 distinct processes, Customer Identification, Due Diligence and Monitoring. As the number of digitized financial services increases, so do the risks for fintech companies and their customers.

Read full article on datafloq.com

Empowering Cyber Defenders: The Role of AI in Securing Our Digital Future

In recent times, AI technologies have transformed the realm of cybersecurity, endowing it with advanced capabilities to detect, prevent, and respond to cyber threats. These technologies use machine learning, natural language processing, and deep learning to analyze data and identify potential threats. Applications: Machine learning is used in cybersecurity to detect threats, classify them, and identify patterns to prevent future attacks. Applications: NLP is used in cybersecurity to analyze human language-based data such as emails and social media to identify potential threats.

Read full article on datafloq.com



Revolutionary Neuromorphic Computing: The Key to Hyper-Realistic Generative AI

Neuromorphic computing is a new field of artificial intelligence (AI) that aims to mimic the structure and function of the human brain. While there are challenges to overcome in developing neuromorphic computing, such as the complexity of the human brain and ethical concerns, the potential benefits of this technology are too great to ignore. Challenges in developing neuromorphic computing systems include data privacy, ethical concerns, and the need for regulatory frameworks to ensure the responsible and beneficial use of these technologies.

Read full article on datafloq.com

The Business Case for Investing in Application Security Testing

A robust security assurance framework can help a business enterprise mitigate risks, prevent data breaches, protect sensitive data and information, comply with regulatory standards, and maintain the stability, integrity, and accessibility of critical systems. With application security testing, such businesses can ensure compliance with industry-specific regulatory norms. By employing professional security testing services, businesses can gain the trust of customers and improve their revenue streams. Investing in application security testing can go a long way in helping businesses establish trust with customers and other stakeholders, protect critical assets, mitigate risks, meet compliance requirements, safeguard reputation, and achieve success.

Read full article on datafloq.com

Unlocking the Power of NFTs: Top 10 Companies Leading the Way in Gaming Innovation in 2023

These firms typically offer a range of services, including platform development, smart contract development, NFT creation, and integration with other blockchain-based systems. Antier is a blockchain development company that specializes in providing unique NFT gaming platforms that are blended according to your needs and requirements. Appdupe is a blockchain development company that specializes in creating NFT gaming platforms that look incredible and can attract gamers from around the world. Mobiloitte is a blockchain development company that specializes in creating NFT gaming platforms based on the needs of its clients.

Read full article on datafloq.com

Data Science Bows Before Prompt Engineering and Few Shot Learning

By tailoring GPT-3.5 with prompt engineering and few shot learning, Common tasks dont require a data scientist, Gupta pointed out. The example of this single shot learning use case would train the system, then by giving the pattern to the model you can ask any question based on that pattern, Gupta concluded. Although such an example may seem trivial, it attests to the ease of use, lack of specialized knowledge, and dearth of technical skills required to tune GPT-3.5 for almost any natural language technology task.

Read full article on insidebigdata.com

4 Ways to Write Data To Parquet With Python: A Comparison

Learn How To Efficiently Write Data To Parquet Format Using Pandas, FastParquet, PyArrow or PySpark. In this article, I will demonstrate how to write data to Parquet files in Python using four different libraries: Pandas, FastParquet, PyArrow, and PySpark. The code below, is a valid example of how to write large datasets to a parquet file, in batches, by combining the power of pandas and fastparquet : In this article, we have explored four different Python libraries that allow you to write data to Parquet files, including Pandas, FastParquet, PyArrow, and PySpark.

Read full article on towardsdatascience.com

Effective Data Visualization: 9 Valuable Tips to Increase the Quality of Your Charts

Take a look at the following two representations: Image by the Author While both are based on the same data, currency conversion is much better represented in form of a table showing the precise numbers rather than a chart, leaving the reader guessing those decimals. In order to avoid confusion and ensure that the reader gets the most out of your visual, always make sure that the axes are labeled properly showing both the dimensions being presented as well as their units and that they are legible.

Read full article on towardsdatascience.com

Using MLflow with ATOM to track all your machine learning experiments without additional code

Using MLflow with ATOM to track all your machine learning experiments without additional code Start storing models, parameters, pipelines, data and plots changing only one parameter Photo by Hans Reniers on Unsplash Introduction The MLflow Tracking component is an API and UI for logging parameters, code versions, metrics, and output files when running your machine learning experiments and for later visualizing the results. In this story, well explain how to use the ATOM library to easily track your models, parameters, pipelines, data and plots.

Read full article on towardsdatascience.com

Tabu Search Simply Explained

An intuitive explanation of the Tabu Search optimization algorithm and how to apply it to the traveling salesman problem Its other main key ingredient is that it prevents the algorithm from visiting previously observed solutions using memory structures to wider explore the search space. Due to the intractability of the TSP for a certain number of cities, we need to resort to heuristics, such as Tabu Search and Simulated Annealing, to provide sufficient solutions in a reasonable amount of time. Lets first list out some pseudo-code on how we will implement the Tabu Search for the TSP:

Read full article on towardsdatascience.com

Visual ChatGPT brings AI image generation to the popular chatbot

Visual ChatGPT is a new model that combines ChatGPT and VFMs, including Transformers, ControlNet, and Stable Diffusion. Visual ChatGPT is a new model that combines ChatGPT with VFMs like Transformers, ControlNet, and Stable Diffusion. Researchers at Microsoft have developed a system called Visual ChatGPT that features numerous visual foundation models and graphical user interfaces for interacting with ChatGPT. With the use of tools like Visual ChatGPT, the learning curve for text-to-image models may be lowered, and different AI programs can communicate with one another.

Read full article on dataconomy.com

Creating an artificial intelligence 101

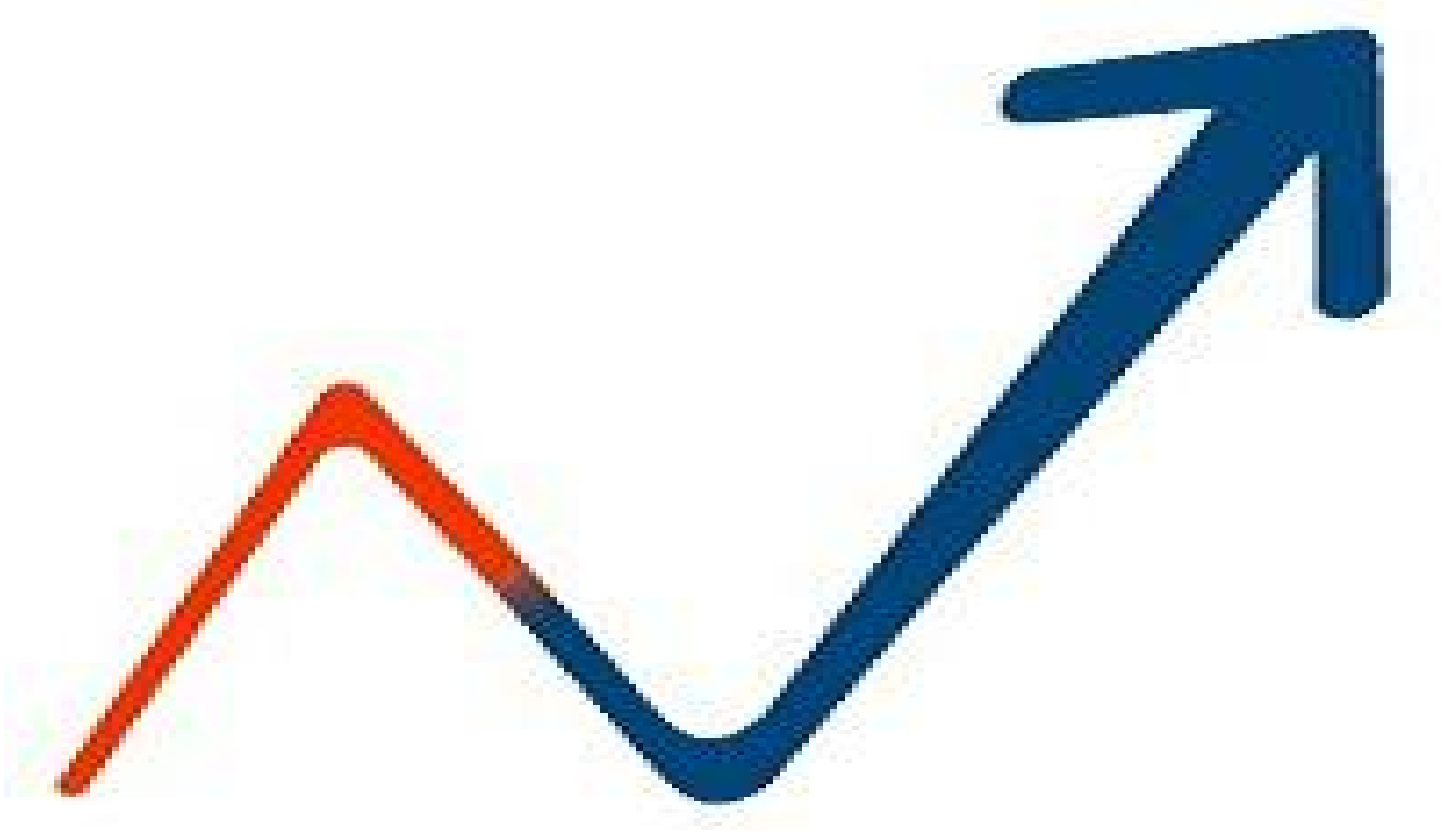
In this article, we will explore the essential steps involved in creating AI and the tools and techniques required to build robust and reliable AI systems. TensorFlow: TensorFlow is an open-source framework developed by Google that provides a range of tools for building and training machine learning models. TensorFlow is an open-source framework developed by Google that provides a range of tools for building and training machine learning models. PyTorch is an open-source framework developed by Facebook that provides a range of tools for building and training machine learning models.

Read full article on dataconomy.com

Level up with Databricks at Game Developers Conference

At Databricks, we specialize in helping game studios build better games with stronger, more engaged communities by empowering teams with a unified platform that seamlessly brings together data, analytics, and AI. We'll also have a presence on the expo show floor, Booth S238 in South Hall, where you can talk directly to data and AI experts, walk through demos demonstrating game analytics and advanced data use cases, and learn how Databricks can help you harness the power of game data to differentiate and drive success for your title like never before.

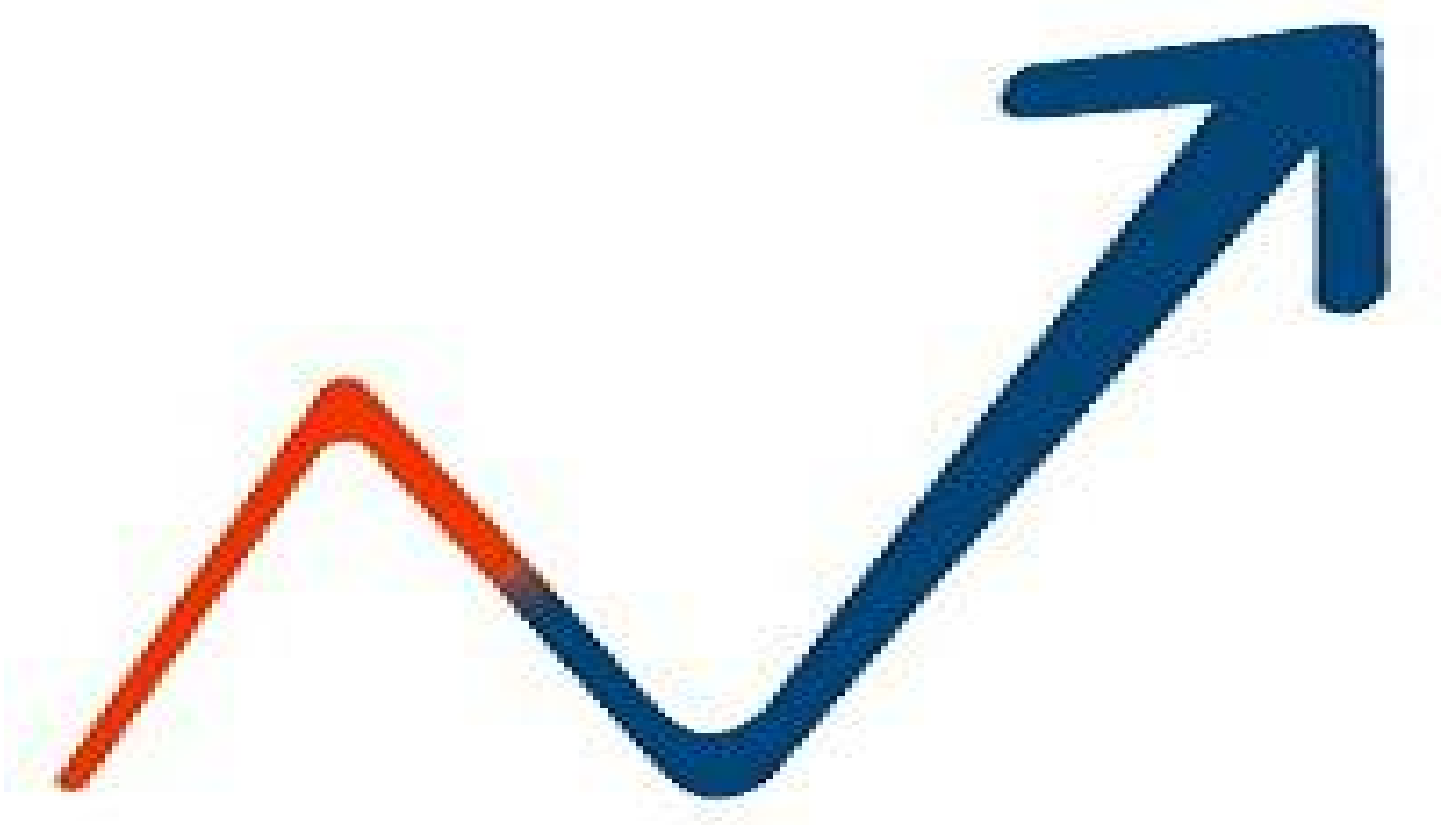
Read full article on www.databricks.com



Mediapipe Tasks API and its Implementation in Projects

This article will explore three exciting projects using the Mediapipe Tasks API focused on a separate domain: Audio, Image, and Text. With the recent release of the Mediapipe Tasks API, developers can now access pre-trained Deep Learning models for various tasks, including Audio, Image, and Text processing. Understand how Mediapipe Tasks API can simplify the process of object detection by providing pre-trained models and machine learning algorithms. #use Mediapipe Tasks API base_options = python.BaseOptions(model_asset_path=model) options = vision.ObjectDetectorOptions(base_options=base_options,score_threshold=0.5) detector = vision.ObjectDetector.create_from_options(options) #using Mediapipe Image Attribute initialize the input image path.

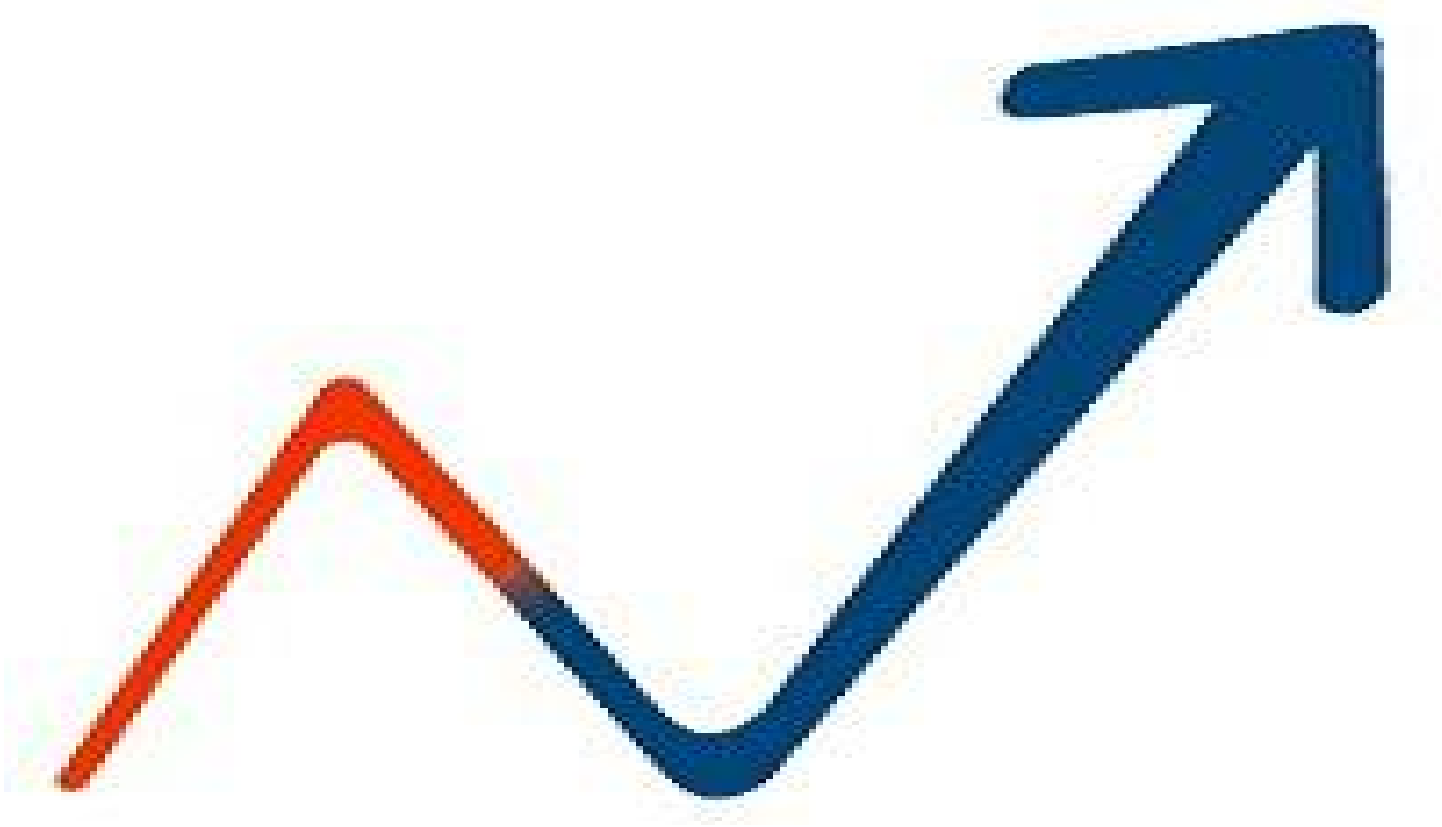
Read full article on www.analyticsvidhya.com



An Introduction to Large Language Models (LLMs)

Lets look into how Hugging Face APIs can help generate text using LLMs like Bloom, Roberta-base, etc. In addition to accelerating natural language processing applications like translation, chatbots, and AI assistants large language models are used in healthcare, software development, and use cases in many other fields.'" output = query({ 'inputs': full_text, 'parameters': params }) pprint(output) Large Language Models (LLMs) have revolutionized the field of natural language processing, allowing for new advancements in text generation and understanding.

Read full article on www.analyticsvidhya.com



Detect Cyberbullying Using Topic Modeling and Sentiment Analysis

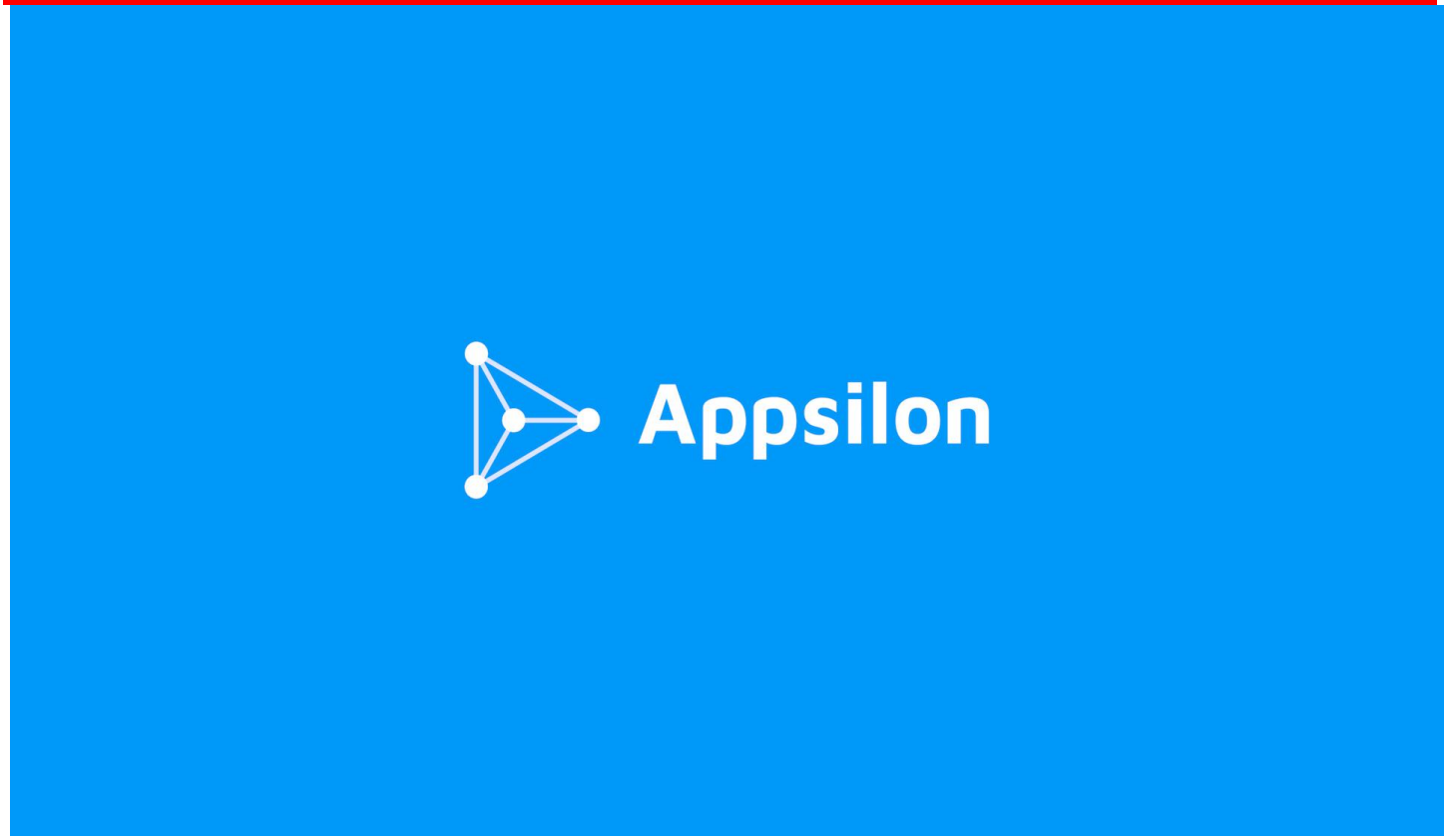
In this article, we will cover an unsupervised learning method of Topic Modeling and a supervised learning method of Sentiment Classification to identify topics in the dataset. This article explores the importance of Topic Modeling for large amounts of text corpus over supervised learning methods with hands-on project implementation. Sentiment classification is an unsupervised machine learning approach to extract frequently discussed topics from a certain text corpus. BERTopic It uses the pre-trained model BERT to embed words in the text corpus to extract topics in large collections of documents.

Read full article on www.analyticsvidhya.com

Advance diversity and inclusion in voice AI with speech disentanglement

Their presentation focused on speech disentanglement and how Amazon uses this technique to influence different aspects of speech tone, phrasing, intonation, expressiveness, and accent to create unique Alexa responses. In this presentation we talked about how we use machine learning (ML) techniques in text-to-speech (TTS) to improve diversity, equity, and inclusion (DEI), to make Alexas response work optimally for everyone. We can use ML techniques to modify various aspects of speech and to improve the diversity and style of TTS voices, thereby addressing the needs of various customers.

Read full article on www.amazon.science



ShinyConf 2022 Short Recap of the 2022 R Shiny Conference (Projects, Talks, and Showcases)

Our Appsilon R Shiny Conference resulted in close to 20 hours of materials covering tech talks, showcases, and projects all spread out over 31 videos. Join Marcin again this year with his talk (03/15/2023): Sharing app state between Shiny modules. Its a combination of a couple of sessions from different Appsilonians discussing everything you need to know to make your Shiny apps stand out. Shiny is an excellent tool for building fast application prototypes, but as your app grows, it will need a proper testing architecture.

Read full article on appsilon.com

If you did not already know

In this work, we develop a distributed model-free, DeepPool, that uses deep Q-network (DQN) techniques to learn optimal dispatch policies by interacting with the environment. Further, DeepPool efficiently incorporates travel demand statistics and deep learning models to manage dispatching vehicles for improved ride sharing services. In this paper, we propose a novel method unsupervised ensemble learning via Ising model approximation (unElisa) that combines a pruning step with a predicting step. We have shown that an IndRNN can be easily regulated to prevent the gradient exploding and vanishing problems while allowing the network to learn long-term dependencies.

Read full article on analytixon.com

Timeseries

Data Processing and Modeling The availability of large quantity of cheap sensors brought forth by the so called Internet of Things has resulted in an explosion of the amounts of time varying data. Understanding how to mine, process and analyze such data will only to become an ever more important skill in any data scientists toolkit. In this lecture, we will work through the entire process of how to analyze and model time series data, how to detect and extract trend and seasonality effects and how to implement the ARIMA class of forecasting models.

Read full article on medium.data4sci.com