Project proposal for the year 2023

Sunday, 24 September 2023 22:09

Proposal sections
- Project description
- Background information
- Scope
- Timeline
- *Budget
- Key outcomes
- Milestones and deliverables
- Risk management
- *Stakeholders and team members

Description
I aim to create a machine learning model to predict trends within the tech community, immerging technologies and hiring practices. The model will leverage historical as well as real time data to identify parterns and trends providing valuable insight into the future of the tech community

Timeline
As the project will be conducted as my fourth year computer science project the aloted time is months with presentations slated for early april. With the first two weeks dedicated to information gathering followed by two more weeks for data cleaning then two more weeks to automate both processes then onwards to model development and model training

Key outcomes
Expected key outcomes are timely and accurate predictions on the future of the tech industry, popularity and usage of different technologies, hiring trends and practices
As well as sorted and formated data regarding tech trends and their proliferation

Scope
The scope of this project includes data collection, preprocessing, model development, testing, and deployment. The model will be trained on relevant technology-related data and job listings and fine-tuned to maximize prediction accuracy.

Milestones
Data collection automation|dataset
Model training|model
Testing and deployment | reports on the ML perfomance

Stakeholders and team members
Tech companies ,market analysts and investors interested in technology trends
New comers entering the tech space looking for confirmation on which stacks to choose
Veterans / tech professionals looking to upskill
Consumers
Industry associations

Datasets
Twitter and reddit sentimental analysis dataset
Stack exchange API
Web traffic time series
Five thirty eight dataset*
Github
News and stock exchange data
Worldbank data catalogs

Project Name: TrendLens
Presenter Name: Collins Iregi

Description
In the ever-evolving landscape of the technology community, our primary aim is to construct a machine learning model that can anticipate and track trends. This includes the emergence of new technologies and shifts in hiring practices.

To achieve this, we plan to leverage a diverse range of data sources. These include public datasets such as the Stack Exchange API, web traffic time series, GitHub trending repositories, and World Bank data catalogs. We also intend to utilize sentiment analysis datasets from platforms like Twitter and Reddit. To capture the pulse of real-time trends, we will employ web crawling techniques on job boards, tech blogs, and Twitter threads.

The model will be meticulously designed to analyze these data sources effectively. It will identify patterns and trends that can provide valuable insights into the future trajectory of the tech community.

We believe this project holds significant potential for impact within the tech community. It aims to benefit various stakeholders including job seekers, learners, market analysts, and consumers by providing them with predictive insights into technology trends.

This proposal outlines a comprehensive plan for a project that combines robust data sources with advanced machine learning techniques to track and predict key trends in the technology community.

Background
Context
With the rapid pace of technological advancement, staying ahead of trends is crucial for businesses and individuals alike. Our model aims to provide a solution by predicting these trends before they become mainstream.

Problem statement
Trends in the technology field are hard to predict as a result of the dynamic nature of the field. The dynamic nature of the tech field, characterized by rapid technological advancements and shifting market demands, makes predicting trends a complex task.

Significance
The significance of this project lies in the value it holds most of all to young professionals and students. The project will merely serve as a guide validating their choices in tech fields or tech stacks by offering a data supported estimation on their longevity and virality.
The project also stands to be of great use to educators as they plan out their curriculums, offering insight on how to tailor their lessons and projects to more adequately prepare their students for the dynamic and ever-evolving tech industry. By aligning educational content with current and predicted tech trends, educators can ensure their students are equipped with relevant skills and knowledge that will remain valuable in the future.
This project also stands to be of value to market analysts and investors. Working in tandem with other forms of market research the project can offer additional insight to the viability of a technological trend as worthwhile investment .

Project Scope

- Data collection and preprocessing- this involves gathering data from multiple public sources and processing it into a format that will be usable by our model
- Model development testing and deployment - this will involve utilizing multiple machine learning techniques such as Artificial Neural Networks(ANN),Long Short-Term Memory(LSTM) and Support Vector Machines(SVM) to better classify the data, find patterns and predict trends.
- The scope of the project's data will be limited to only publicly available sources as a consequence of budget constraints
- The models predictions will mainly focus on career and career growth opportunities this is in a bid to limit the complexity of the model

Resources

- UCI machine learning repository
- Stack exchange API
- Web traffic time series by Google
- Support & outside services
- GitHub
- Data from web scrapping twitter threads, reddit forums and tech blogs

Key outcomes
The success of this project will be measured by the following metrics:

- Delivery of a machine learning model whose success will be based on its accuracy
- Delivery of a data acquisition tool whose success will be measured by the efficiency of data collection and cleaning, timeliness of the data and amount and accuracy of the data

Milestones and deliverables
The entire project is slated to last six months with that time being broken down into a series of milestones each with a specific deliverable. They are as follows:

- Data collection with the deliverable being precise and concise datasets at regular timely intervals- this is slated to occupy the first three weeks with data collection and submission to follow on a weekly interval
- Model training with the deliverable being a fully functioning model judged by basis of its accuracy- this is slated to take four months allowing ample time to test and implement multiple combinations of machine learning techniques
- Testing and implementation that will deliver reports on the overall accuracy of the model and success of the project - this is slated to take a month. The reports will be a measure of the project's success and contribute significantly to the project's final presentation.

Risk management
Potential issues might include:
- Data quality issues - these however can be mitigated by rigorous data validation practices
- Timeline delays- this can be alleviated by the allocation of buffer time within the schedule
- Inaccuracies as a result of economic conditions

Stakeholders
Some of the stakeholders who stand to benefit from this project are:

- Tech companies, market analysts and investors
- Students as well as professionals looking to join the tech space
- Educators planning out their curriculums
- IT professionals looking to upskill
- Tech consumers