ExEEd-PROJECT BASED LEARNING
REPORT
ON

Machine Learning model for Product risk management

A Report submitted

by

B.Chakradhar 21951A1220
B.Datta Sai 21951A1222
K.Jaya Krishna 21951A1233

Information Technology
INSTITUTE OFAERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad–500 043, Telangana

Problem Definition:

Traditional risk management approaches often fall short in comprehensively analyzing the diverse and unstructured textual data available from sources such as customer reviews, social media, and industry reports. This requires an innovative solution that harnesses the power of Natural Language Processing (NLP) with machine learning techniques to enhance the precision and depth of product risk management.

Developing a machine learning model for product risk management involves addressing various challenges related to identifying and mitigating risks associated with products such as:
1.Risk Identification
2.Integration with Existing Systems
3.Scalability
4.Toxic comment identification

Gathering Requirements:

1.Data Sources:

Identifying and gather information on the relevant data sources for the NLP model. This includes customer reviews, social media data, industry reports, and other textual data related to product feedback and market trends.

2.Stakeholder Identification:

Identify key stakeholders involved in the product risk management process, including product managers, data scientists, business analysts, and end-users.

3.Data Preprocessing:

Define the preprocessing steps required for the textual data. This involves text cleaning, tokenization, and the removal of stop words and irrelevant information. Determine how to handle missing or incomplete data.

4.Feature Selection:

Determine the key features or attributes to be extracted from the textual data for analysis. Specify whether sentiment analysis, topic modeling, or other NLP techniques will be utilized and why.

5.Model Training and Evaluation:

Define the machine learning algorithms to be employed for risk prediction. Establish the criteria for evaluating the model's performance, such as precision, recall, and F1 score. Decide on the training and testing datasets.

6.Business Objectives:

Clearly define the overarching business objectives for implementing the machine learning model. This may include improving risk assessment accuracy, reducing time-to-market, and enhancing overall product success rates.

7.Integration with Existing Systems:

Assess the compatibility and integration requirements with existing product management systems, databases, or analytics platforms. Ensure seamless communication between the NLP model and other tools in use.

8.Documentation and Training:

Plan for comprehensive documentation of the model, including its architecture, algorithms used such as Logistic Regression, Naive Bayes, Support Vector Machines, and user guidelines.
Use a model trained of a large corpus of data. Transformer model accounts for the words but also the context related to other words.

Design:

Designing a machine learning model for product risk management using Natural Language Processing (NLP) involves defining the architecture, workflow, and components of the system. Below is an outline for the design process:

System Architecture:

Input Data: Collect and integrate data sources including customer reviews, social media comments, and industry reports.
Preprocessing Module: Implement tokenization, lemmatization, and stop-word removal using NLP libraries (e.g., spaCy).
Feature Extraction using NLP: Utilize NLP techniques such as sentiment

analysis, topic modeling, and named entity recognition.
Labeling Module: Define labels based on identified risks for supervised learning.

Model Architecture:

CNN Model:
The system is trained using the CNN model which consists of four layers, i.e.,
1.Input layer: The input layer is where the network receives the initial data, and for image data, it is typically a multi-dimensional array representing the pixel values of the image.
2.Convolution layer: Fundamental building block responsible for extracting features from input data
3.Global max pool layer: Global Max Pooling is a pooling layer that takes the maximum value from each feature map.
4.Fully connected layer: The fully connected layer processes the features extracted by the convolutional layers and makes predictions.

Monitoring and Maintenance:
Implement tools for monitoring the model's performance and also set up alerts for potential issues.
Establish a regular maintenance schedule for updates and improvements.

Feedback Loop:
Establish a feedback loop for gathering user feedback. Use feedback to inform model improvements and updates.

Implementation:

Data Collection:

Gather relevant data sources, including customer reviews, social media comments, industry reports, and any other textual data related to product feedback and market trends. Ensure the dataset is representative of the various stages of the product lifecycle.

Data Preprocessing:

Clean and preprocess the textual data. This may involve tasks such as:
Tokenization: Breaking down text into individual words or phrases.
Removing stop words: Eliminating common words that do not contribute to the overall meaning.
Lemmatization or stemming: Reducing words to their base or root form.

Handling missing or incomplete data.

Feature Extraction using NLP:

Utilize NLP techniques to extract relevant features from the preprocessed text. Common NLP tasks include:
Sentiment Analysis: Determining the sentiment (positive, negative, neutral) expressed in the text.
Topic Modeling: Identifying key topics or themes within the textual data.
Named Entity Recognition (NER): Extracting entities such as product names, organizations, and locations.

Labeling Data:

Define and label the data based on the identified risks. For supervised learning, this involves creating a labeled dataset with examples of different risk categories.

Model Selection:

Choose appropriate machine learning algorithms for the task at hand. Common models for NLP tasks include:
Natural Language Processing Libraries: Such as spaCy, NLTK, or Hugging Face Transformers.

Machine Learning Models: Support Vector Machines (SVM), Random Forest, or neural network architectures like LSTM or BERT.

Model Training:

Split the labeled dataset into training and testing sets. Design a Convolutional Neural Network for text classification. Include layers for embedding, convolution, and global max-pooling. Use a dense layer for binary classification with a sigmoid activation function.

Evaluation Metrics:
Define appropriate evaluation metrics for toxic comment identification.
Common metrics include accuracy, precision, recall, F1 score, and area under the ROC curve (AUC-ROC).

Real-time Monitoring Integration:

If real-time monitoring is a requirement, integrate mechanisms for continuous learning and adaptation to evolving data patterns. This may involve periodic model retraining or online learning strategies.

Security and Compliance:
Implement security measures to protect sensitive information in the comments.
Ensure compliance with privacy regulations and guidelines.

Model Deployment:

Deploy the trained model into the product risk management system. Ensure seamless integration with existing systems, databases, or analytics platforms.

Feedback Loop Implementation:

Establish a feedback loop for ongoing model improvement. Gather feedback from users and monitor the model's performance in real-world scenarios to inform future iterations.

CONCLUSION

Machine learning have the potential to transform product risk management in industrial services. By analyzing large volumes of data and identifying patterns, these technologies can help businesses to identify potential risks and take proactive measures to mitigate those risks. It not only allows for real-time monitoring, continuous learning, and adaptability to evolving risks but also provides a mechanism for the systematic identification and removal of toxic comments. The model becomes an integral part of the product management ecosystem, contributing to a safer and more user-friendly environment.
The future of machine learning for product risk management is dynamic, with continuous advancements in technology, increased data availability, and growing awareness of the benefits of AI in enhancing safety, quality, and compliance across various industries. As organizations adapt to these developments, they will be better equipped to identify and mitigate risks, ultimately improving product safety and consumer satisfaction.
The accuracy of the model is 71 -75%.

INSTITUTION'S INNOVATION COUNCIL
MOE'S INNOVATION CELL

Institute Name:
Institute of Aeronautical Engineering

Title of the Innovation/Prototype:
A MACHINE LEARNING MODEL FOR PRODUCT RISK MANAGEMENT

Team Lead Name:	Team Lead Email:	Team Lead Phone:	Team Lead Gender:
Chakradhar Butharasi	21951a1220@iare.ac.in	9542532209	Male

FY of Development:	Developed as part of:	Innovation Type:	TRL LEVEL:
2023-24	Academic Research Assignment/Industry Sponsored Project	Service	1

Theme:
Other Emerging areas Innovation for Start-up,

Define the problem and its relevance to today's market / sociaty / industry need:
This model helps to identify the risk of product manufacturing and marketing.

Describe the Solution / Proposed / Developed:
A machine learning model is a program that can find patterns or make decisions from a previously unseen dataset. For example, in natural language processing, machine learning models can parse and correctly recognize the intent behind previously unheard sentences or combinations of words. This model helps to identify the risk of product manufacturing and marketing.

Explain the uniqueness and distinctive features of the (product / process / service) solution:
Machine learning has the potential to transform product risk management in industrial services. By analyzing large volumes of data and identifying patterns, these technologies can help businesses to identify potential risks and take proactive measures to mitigate those risks.

How your proposed / developed (product / process / service) solution is different from similiar kind of product by the competitors if any:
The experiments are conducted using different numbers of convolution layers with varying numbers and sizes of filters. The CNN models are trained on 50% of the dataset and tested on the remaining 50% of the dataset. Our model can achieve better performance than traditional ML approaches and has achieved an accuracy of 95%.

Is there any IP or Patentable Component associated with the Solution?:
No

Has the Solution Received any Innovation Grant/Seefund Support?:
No

Are there any Recognitions (National/International) Obtained by the Solution?:
No

***Is the Solution Commercialized either through Technology Transfer or Enterprise Development/Startup?:**
No

Had the Solution Received any Pre-Incubation/Incubation Support?:
No

Video URL:
https://drive.google.com/file/d/10kBxSN8EvTUtHZdnsw4uDvM--LK55tiR/view?usp=sharing

Innovation Photograph:
[View File]