VISVESVARAYA TECHNOLOGICAL UNIVERSITY
JNANASANGAMA, BELAGAVI – 590018

An Internship Report
on

MACHINE LEARNING WITH AI USING PYTHON

Submitted in partial fulfillment for the award of degree of

Bachelor of Engineering
in

Computer Science and Engineering

Submitted by

FERRIN MARIA CHRISTINA DSOUZA

4SO17CS038

Internship Carried Out
at
Knowledge Solutions India
Ghanshyam Park Society, Dhole Patil Rd, Pune, Maharashtra-411001

Internal Guide
Mr. Gerald Harry Fernandes
Assistant Professor
St Joseph Engineering College

External Guide
Mr Gurvansh Singh
Project Mentor
Knowledge Solutions India

Department of Computer Science and Engineering
St Joseph Engineering College
Mangaluru - 575028

2020-2021

VISVESVARAYA TECHNOLOGICAL UNIVERSITY
JNANASANGAMA, BELAGAVI – 590018

An Internship Report
on

MACHINE LEARNING WITH AI USING PYTHON

Submitted in partial fulfillment of the requirements for the degree

Bachelor of Engineering
in

Computer Science and Engineering

Submitted by

FERRIN MARIA CHRISTINA DSOUZA **4SO17CS038**

Department of Computer Science and Engineering
St Joseph Engineering College
Mangaluru - 575028

2020-21

St Joseph Engineering College
Mangaluru – 575 028
Department of Computer Science and Engineering

CERTIFICATE

Certified that the Internship Work titled Machine Learning with AI using Python was carried out by Ms. Ferrin Maria Christina Dsouza, bearing USN 4SO17CS038, a bonafide student of final year B.E. inpartial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering of the Visvesvaraya Technological University, Belagavi, during the year 2020-2021. Further, it is certified that all corrections/suggestions indicated during Internal Evaluation have been incorporated in this report.

---	---	---
Mr Gerald Harry Fernandes	**Dr Sridevi Saralaya**	**Dr Rio D'Souza**
Internal Guide	**Head of the Department**	**Principal**

External Viva Voce Examination

Name of the Examiners	**Signature with Date**
1.	---
2.	---

Microsoft Partner

Knowledge Solutions India
Skill development | Certification | Placement prep

Ref KSI-ML-1506-2707-317

Certificate of Internship

This is to certify that

FERRIN MARIA CHRISTINA DSOUZA

of **ST.JOSEPH ENGINEERING COLLEGE , MANGALORE**

Has successfully undergone a summer training & internship of 6 weeks on

MACHINE LEARNING WITH AI USING PYTHON

Along with project on

RESTAURANT REVIEWS

From **15.06.2020** to **27.07.2020** with Knowledge Solutions India

Date of Issue
14/8/2020

KL
Authorized Signatory
Kumar lala

Certificate of COMPLETION

Course Completion

This is to certify that

FERRIN MARIA CHRISTINA DSOUZA

Enrollment id - **SIT-ML-353**

Has successfully completed Course in **MACHINE LEARNING WITH AI USING PYTHON**

15/7/2020
Date

Signature

Knowledge Solutions India
Skill development | Certification | Placement prep

Microsoft Partner Network

CERTIPORT
AUTHORIZED TESTING CENTER

Microsoft AEP Authorized Education Partner

DECLARATION

I, **Ferrin Maria Christina Dsouza**, bearing **USN 4SO17CS038**, student of final year B.E. in Computer Science and Engineering, St Joseph Engineering College, Mangaluru, hereby declare that the Internship Work titled**"MACHINE LEARNING WITH AI USING PYTHON"** has been duly executed by me from June to July 2020, at Knowledge Solutions India (KSI), Maharashtra. Further, the "Tasks Performed" section of this report represents the work done solely by me and does not contain any statements falsely claiming work done by others, as my own.

Date:

Place: Mangaluru **Ferrin Maria Christina Dsouza**

ACKNOWLEDGEMENT

We dedicate this page to acknowledge and thank those responsible for the shaping of the project. Without their guidance and help, the experience while constructing the dissertation would not have been so smooth and efficient.

We extend our heartfelt gratitude to **Mr. Gurvansh Singh** of **Knowledge Solutions India**, for his guidance, constant supervision as well as providing us the necessary information regarding the project. Without his willing disposition, the spirit of accommodation, frankness, timely clarification and above all faith in us, this project could not have been completed in due time. His readiness to discuss all important matters at work deserves special attention.

We sincerely thank **Mr. Gerald Harry Fernandes, Assistant Professor**, Department of Computer Science and Engineering, for his guidance and valuable suggestions which helped us to complete our internship project.

We owe our profound gratitude to**, Dr Sridevi Saralaya, Head of the Department**, Computer Science and Engineering, whose kind consent and guidance helped us to complete this work successfully.
We are extremely thankful to our Principal **Dr Rio D'Souza,** our Director **Rev Fr Wilfred Prakash D'Souza and** our Assistant Directors **Rev Fr Rohith D'Costa** and **Rev Fr Alwyn Richard D'Souza** for their support and encouragement.

We also like to extend thanks to our friends and family members for their continuous support.

EXECUTIVE SUMMARY

I carried out my internship as a Machine Learning Intern at Knowledge Solutions India (KSI) from 15th June 2020 to 27th July 2020.

KSI is a certification and training company. Being Microsoft Authorised Education Partners as well as Certiport CATC they offer International Certifications by Microsoft, Apple, Adobe, EC Council, Autodesk, Quick books etc. And work closely with Universities and Colleges across the country. KSI has a well-qualified team of subject matter experts. These professionals have an abundance of experience in their subjective fields and are also certified themselves. They are passionate about the subjects they teach and bring this enthusiasm into their webinars and courses.

The internship's goal was to develop a Machine learning model using Natural Language Processing on Sentiment Analysis of Restaurant Reviews. The purpose of this analysis is to build a prediction model to predict whether a review on the restaurant is positive or negative. To do so, we worked on Restaurant Review dataset, we loaded it into predictive algorithms Multinomial Naive Bayes, Bernoulli Naive Bayes and Logistic Regression. In the end, we hoped to find a "best" model for predicting the review's sentiment. To build a model to predict if review is positive or negative, following steps such as Importing Dataset, Preprocessing Dataset, Vectorization, Training and Classification, Calculating the accuracy of the models, Analysis Conclusion were performed.
The internship gave me an opportunity to develop my professional skills and to obtain experience in the industry. It also inspired me to research on various optimum machine learning models and how to tackle them to meet real world problems. Additionally, it helped me develop my skills in teamwork, communication and time management.

I was able to complete my assigned tasks within the prescribed time frame. This internship also gave me the opportunity to apply the knowledge acquired to actual work experiences. The internship overall helped me gain valuable work experience and improved my skills.

Ferrin Maria Christina Dsouza

CONTENTS

FIGURE INDEX

CHAPTER 1

ABOUT THE COMPANY

1.1 Brief History

KSI is a Delhi based online skilling and certification company and a Microsoft Authorized Education Partner. The company was incorporated on 15 March 2006 in Pune, Maharashtra. KSI is a partner company of Quantum Learnings Group. KSI employs approximately 50 people in full-time positions. Mr. Sandeep Jethani, the Founder of KSI is an innovator, entrepreneur, ed-tech educator.

The company is India's private sector enterprise, with partnership businesses and is an e-learning platform. The services provided by them are certification, skill development, placement preparation and training. It's headquartered is in Maharashtra. Being Microsoft Authorized Education Partners as well as Certiport CATC, they offer international certifications by Microsoft, Apple, Adobe, EC Council, Autodesk, Quick books etc. and work closely with universities and colleges across the country. KSI has a well-qualified team of subject matter experts. These professionals have an abundance of experience in their subjective fields and are also certified themselves. They are passionate about the subjects they teach and bring this enthusiasm into their webinars and their courses.

1.2 Major Milestones

KSI was founded in the year 2006 to help students to get trained, certified and undertake internships. The company also makes sure that every approach produced at KSI operates effectively and delivers tangible results to the customers.

1000 plus engineering projects were built by the organization in various fields such as machine learning, data science, IOT, cyber security, ethical hacking and much more. Mentorrbuddy was founded in May 2016 by Sandeep Jethani. It is a platform introduced to improve the overall learning experience at KSI and even can get help for placement preparation using this platform. Quantum Learning was founded in March 2020. Quantum Learning was introduced to help upskill and earn certifications in data science, data analytics, machine learning and business analytics.

1.3 Services Offered by the Company

The company provides technology and robotics laboratory that is specifically designed to uncover rationale behind technical wonders for young minds aged 12 + and constructs both small business and corporate e-commerce websites. There is a dedicated team of tech professionals designing android apps for musicians, writers, businesses, as well as personalized applications, and e-commerce applications. The firm is adapting existing Internet and Windows Application development techniques to offer custom software services to simplify consumer market. In KSI Micro level performance monitoring is available through the subject and topic-wise analytics which provide detailed performance analytics on strengths and weaknesses.

The company even helps to understand the topics better by providing access to more than 10,000 learning videos, including concept and solution videos.

KSI trains students in the field of Mobile app development, Progressive Web app development, Angular Js, Python, IoT, ML, AI, Data Science, Adv Excel or Digital Marketing. It has the best technical training delivered across the country. All of their courses start from scratch assuming no prerequisite and the team puts full efforts to ensure that a candidate completed the program only after acquiring relevant skills.

KSI has a platform known as Mentorrbuddy to improve the learning experience at their company. It is an online learning and evaluation platform with AI and ML capabilities. It is a unique system designed, with intricately integrated AI and Machine Learning algorithms which provide you with detailed analysis after every test and helps you rectify your weak areas and work on them. They have company specific tests and content to help you narrow down your search and practice on areas which are more important according to your goals. Mentorrbuddy helps you to understand the concepts easily during internship, and boosts the aptitude skills.

The KSI summer internship program is a cycle consisting of five processes. The first process is project allocation to students. The next step after the completion of the previous process is to provide online training on technology related to the project. The following step is certification on technology related to the project. The next process is online presentation by students on the project followed by internship letter and Microsoft certification.

KSI are a part of the prestigious Micosoft Global Training Partner program. The job-oriented programs they offer are for freshers with experience between 0 - 2 years, looking to get into the field of Data Science, Data Analytics, and Business Analytics and up skill themselves. The courses are bundled with Microsoft Certifications and 100% Placement Assistance.

Along with concepts and tools used in Data Science they will also train for placement tests, assignments and interviews to ensure complete development of the student, making them industry ready using Our Placement Assistance tool, MentorrBuddy.

1.4 Contact Details

Company Location: Knowledge Solutions India,
Ghanshyam Park,
Dhole Patil Road,
Pune, Maharastra

Mobile Number: +91 - 9910732501 / 9818864887

Email Address: rajkumarbhunia.ksi@gmail.com

CHAPTER 2

ABOUT THE DEPARTMENT

2.1 Introduction

KSI is a software development and educational training company in Pune, Maharashtra. It was nurtured by a group of entrepreneurs with a sole mission of establishing a dedicated Research & Development Cell to fertilize innovative technologies. Their team always ensures that every solution that is developed is efficient in working and brings tangible results to their clients.

KSI has developed 1000 plus engineering projects in various fields viz. Machine Learning, Data Science, IOT, Cyber Security, Ethical Hacking and many more. As an intern, we were taught about various machine learning algorithms and how these can be used to solve real world problems.

2.2 Roles and responsibilities

- ML and AI requirements including all the processes from data collection, cleaning, and pre-processing, to training models and deploying them to production.
- Passion towards artificial intelligence and stay up-to-date with the latest developments in the field.
- Understanding business objectives and developing models that help to achieve them, along with metrics to track their progress.
- Analysing the ML algorithms that could be used to solve a given problem and ranking them by their success probability.
- Exploring and visualizing data to gain an understanding of it, then identifying differences in data distribution that could affect performance when deploying the model in the real world.
- Verifying data quality, and/or ensuring it via data cleaning.
- Supervising the data acquisition process if more data is needed.
- Finding available datasets online that could be used for training.
- Defining the pre-processing or feature engineering to be done on a given dataset.
- Training models and tuning their hyperparameters.
- Analysing the errors of the model and designing strategies to overcome them.

CHAPTER 3

TASKS PERFORMED

3.1 Weekly Work Plan

Week 1:

- On the first week, we were introduced to Machine Learning and its applications.
- We were taught different regression model and their goal in forecasting, predicting or error reduction to fit a predictive model to observed data set.
- We were assigned tasks on the above topics, both hands-on as well as to conduct further research about them.
- We were asked to implement data pre-processing methods such as EDA, extract x and y from dataset, handling missing data etc.
- We assigned the task to implement Linear Regression and Polynomial Regression and visualize output.
- At the end of the week we were asked to develop a code for the automatic implementation of backward elimination for the dataset.

Week 2:

- We were introduced to some of the statistical measures such as Entropy, Gini impurity and variance.
- Machine Learning models such as Decision tree, random forest was also taught and their use in classifying a ML problem based on the given dataset.
- We were also taught about the standard measures that can be used to summarize how good a set of predictions can be such as mse, rmse and accuracy, confusion matrix and r2_score. We had to apply the predictions to calculate the mse, rmse and r2_score for different value of n estimator.
- We were assigned a task to create a Random Forest Model on a given dataset and calculate the mse, rmse, r2_score of our Random Forest model with 10 estimators. We were also asked to create and train the Decision tree Model and visualize its output and also plot the values by increasing the precision.
- We were taught some of the very powerful ensemble methods such as bagging, boosting and stacking and how they will be used in Random forest model.
- Some of the key parameters for decision tree modelling such as tree pruning, grid search technique was also taught.
- We were then assigned tasks on the above topics both hands-on as well as to conduct further research about them.

Week 3:

- We were taught two of the Supervised machine Learning models such as SVM, SVM-PCA and KNN and their application in pattern recognition problem in medicine.
- We were also taught that linear support vector machines are parametric in nature, in order for SVM to capture non parametric pattern we need to use kernels such as RBF, polynomial etc.
- Task was given to analyse the behaviour and accuracy using different SVM kernels, poly component analysis of SVM model.
- Notes on different distance matrices used in KNN algorithm such as Euclidean Distance, Hamming Distance, Minkowski distance was explained.
- We were assigned tasks on the above topics both hands-on as well as to conduct further research about them.

Week 4:

- In the 4th week we were taught about logistic regression, multilinear logistic regression and K-means clustering.
- We were assigned the task to create and train K-Means Model and visualize the cluster.
- We were also taught differences between logistic and linear regression and their uses.
- Different types of clustering technique such as fuzzy, density etc. was also taught.
- We were assigned tasks on the above topics both hands-on as well as to conduct further research about them.
- We were also assigned with the task to work with same K-means but with different column combinations.

Week 5:

- We were taught about artificial neural network and back propagation.
- Deep Neural Network and their power to solve real world problems was also touched upon.
- We were assigned a task to identify different animals through deep neural network by training the model with the dataset given.

Week 6- Project Development:

- I was assigned to a group with the project on Natural Language Processing, particularly Sentiment Analysis of Restaurant Reviews which is demonstrated in the next section.
- At the end of the week, we had to present and submit our project on Sentiment Analysis of Restaurant Reviews to our external internship guide.

3.2 Internship Project

Title: Sentiment Analysis of Restaurant Reviews

3.2.1 Description

The purpose of this analysis is to build a prediction model to predict whether a review on the restaurant is positive or negative. This is done using predictive algorithms such as Multinomial Naive Bayes, Bernoulli Naive Bayes and Logistic Regression. Once the accuracy of each of these models is found, we use the model with highest accuracy in our project to predict if the given comments of user's involve positive or negative Sentiments. To build a model to predict if reviewis positive or negative, following steps such as importing dataset, pre-processing dataset, vectorization, training and classification, analysis and conclusion.

3.2.2 Features

- Restaurant_Reviews.csv is a dataset from Kaggle datasets which consists of 1000 reviews on a restaurant.
- To build a model to predict if review is positive or negative, following steps are performed:
    - Importing Dataset
    - Pre-processing Dataset
    - Vectorization
    - Training and Classification
    - Analysis and Conclusion
- Dataset has 10000 rows and 8 columns.
- We have to predict whether a review is "Positive" or "Negative".
- Porter Stemmer method has been used for Stemming.
- Applied many different algorithms LSTM, Bi-directional LSTM, Random Forest Classifier, MultinomialNB, SVM and KNN and chose the one with highest accuracy.
- After performing data cleaning, I converted "Rating" Column, which is actually a numerical column, into the column that has two labels "Positive" and "Negative".
- I considered Rating Above 3 as "Positive" and Below 3 as "Negative".
- The project as around 10000 reviews and the max_features = 9000 which is best to achieve high accuracy.

3.2.3 Software Requirements

Platform: Flask
Deployment: Heroku platform
Data set: Kaggle
Programming language: Python, HTML, CSS
Library: NLTK

3.2.4 Screenshots

Fig 1. Home Page

Fig 2. Prediction Page

```
In [219]: from tensorflow.keras.layers import Bidirectional
          import keras
          from tensorflow.keras.layers import Dropout
          ## Creating model
          embedding_vector_features=40
          model=tensorflow.keras.Sequential()
          model.add(Embedding(voc_size,embedding_vector_features,input_length=sent_length))
          model.add(Dropout(0.3))
          model.add(Bidirectional(LSTM(150)))
          model.add(Dropout(0.3))
          model.add(Dense(1,activation='sigmoid'))
          model.compile(loss='binary_crossentropy',optimizer='adam',metrics=['accuracy'])

In [220]: model.fit(X_train,y_train,validation_data=(X_test,y_test),epochs=5,batch_size=64)

          Train on 7500 samples, validate on 2500 samples
          Epoch 1/5
          7500/7500 [==============================] - 24s 3ms/sample - loss: 0.4506 - accuracy: 0.7997 - val_loss: 0.2628 - val_accuracy: 0.8952
          Epoch 2/5
          7500/7500 [==============================] - 20s 3ms/sample - loss: 0.2333 - accuracy: 0.9065 - val_loss: 0.2500 - val_accuracy: 0.9060
          Epoch 3/5
          7500/7500 [==============================] - 20s 3ms/sample - loss: 0.1949 - accuracy: 0.9281 - val_loss: 0.2491 - val_accuracy: 0.9008
          Epoch 4/5
          7500/7500 [==============================] - 20s 3ms/sample - loss: 0.1672 - accuracy: 0.9389 - val_loss: 0.2610 - val_accuracy: 0.8964
          Epoch 5/5
          7500/7500 [==============================] - 20s 3ms/sample - loss: 0.1487 - accuracy: 0.9475 - val_loss: 0.2686 - val_accuracy: 0.8936
Out[220]: <tensorflow.python.keras.callbacks.History at 0x192f5f62048>
```

Fig 3. Code to train the model

Algorithm	Accuracy
Random Forest	89.28%
MultinomialNB	90.84%
SVM	76.68%
KNN	81.44%
LSTM	87.56%
Bi-Directional LSTM	89.36%

Fig 4. Accuracy obtained using different models

CHAPTER 4

REFLECTION NOTES

4.1 Experience

My internship at KSI has taught me more than I could have imagined. The internship guide,Mr. Gurvansh Singh is a continuous source of inspiration. The guidance provided by him helped me learn new technologies and complete my internship with ease. I gained valuable work experience and the internship also made me realize how to handle the responsibilities and execute the given tasks within the deadlines.

Operating with a team full of enthusiastic and committed students was nice experience. I waswilling to sharpen my personal and management skills apart from my technical knowledge.To me it was beneficial for work in a dedicated team, to exchange knowledge, to learn new tools and new technologies.

During this internship I learned how to perform data pre-processing. I learnt different libraries in Python programming language that support machine learning. I was taught the applications of machine learning and their power to solve real world problems, additionally I was also taught the types of machine learning algorithms including supervised, unsupervised and reinforcement learning and their working. I also came to know about the mathematics and statistics behind the working of ML algorithms including algebra, analytical geometry,vector composition probability and statistics. Natural Language Processing was one of the other applications of ML which was touched upon. I was thought different NLTK libraries and techniques such as stemming, removing of stop words to eventually obtain high accuracyfor the developed model.

As part of my internship, I worked on the project ‘Sentiment Analysis of Restaurant Reviews’. Our guide at the company was friendly and supported my team at every stage of our project and was quick to solve our doubts.

The project came up really well meeting with all the requirements. It was my first experience to work on a project with due solidarity. I faced several difficulties during the project completion, but my team- mates and guide helped me a lot to make it much more successful.

4.2 Technical Outcomes

1. **Importance of Data and Data Preparation in Machine Learning:**
   ML depends heavily on data. It's the most crucial aspect that makes the algorithm training possible. In our internship we were instructed on how crucial data preparation is in learning process. We were taught ways to refine data, articulating the problems early (to know whether the given problem is a regression, classification, ranking or clustering), establish data collection mechanisms, checking the quality of data (tangibility of human error, outliers, data balance etc), formatting data to makeit consistent and reducing the dimensionality of data by attribute sampling, record sampling and aggregation.

2. **Implementing different types of ML algorithm and applications:**
   We were taught the applications of machine learning and their power to solve real world problems, additionally also taught the types of machine learning algorithms including supervised, unsupervised and reinforcement learning and their working.

3. **Understood the mathematics behind Machine Learning:**
   We also came to know about the mathematics and statistics behind the working of ML algorithms including algebra, analytical geometry, vector composition probability and statistics.

4. **Learned the application of Natural Language Processing:**
   Natural Language Processing was one of the other applications of ML which was touched upon. We were thought different NLTK libraries and techniques such as stemming, removing of stop words to eventually obtain high accuracy for thedeveloped model.

4.3 Non-Technical Outcomes

1. **Teamwork:**
   Being able to work with other people in the department is the key. I worked with rest of the team in developing this application. Interns were encouraged to work effectively, being able to work together and help one another.

2. **Confidence:**
   When we believe in ourselves, we are more likely to achieve the goals we set. But self- doubt in work place is still a big issue for many professionals, even when they have the support of their boss or colleagues. It determines how much we achieve in life and because it has a major impact on our happiness. Self-confidence can be defined as our assessment of our own self-worth.

3. **Skills:**

   Main thing internship taught me is the skill required in that field. It taught me how to think about software on the business level rather than a personal level. I learnt how to take company's values, needs and voice and put it in characters.

4. **Organizing the task:**

   Scheduling of the task is important for day-to-day completion of the task. We madeto-do list at the end of each day for the following day. This helped us prioritize the task and make sure that the task is attainable.

5. **Open Mindedness:**

   Every person in the team is a unique individual, with different ideas and mindset. So being a good communicator as well as a listener is necessary. Hence being open to listening and understanding other persons point of view is an important aspect of team work. By being willing to enter into a dialogue. Even with people with who we disagree, we will be able to have more honest, productive conversations.

6. **Clarity and conclusion:**

   Good communications mean saying just enough- not talking too much or too little. We need to learn to try to convey our message in few words as possible. We should tell what we want clearly and directly. We should think about what we want to say before we actual communicate it, to avoid talking excessively or confusing our audience.

CHAPTER 5

REFERENCES

1. https://www.kaggle.com/ -Jupyter Notebooks environment. Access free GPUs and ahuge repository of community published data & code.
2. https://towardsdatascience.com/ - a medium publication sharing concepts, ideas andquotes.
3. “Natural Language Processing and information retrieval” by Tanveer Siddiqui.
4. https://www.geeksforgeeks.org/machine-learning/