**A Major Project Report**

**On**

**PREDICTING LIFE EXPECTANCY USING ML**

**Submitted in partial fulfillment of the**

**Requirements for the award of**

**INTERNSHIP CERTIFICATE**

**IN**

**MACHINE LEARNING**

**Submitted By**

**RIKKA PRANEETH**

**Under the guidance of**

**Mr.CHARAN**

**ACKNOWLEDGEMENT**

I would like to express our sincere thanks to Mr.charan for his motivation and co-

operation for the successful completion of the project.

I would like to Express our sincere thanks to my ML-6 batch friends for their encouragement to achieve the goals of the project

**Praneeth rikka**

**DECLARATION**

This is to Certify that the project work entitles “Predicting life Expectancy” submitted to smartinternz in partial fulfillment of the requirement for the award of Internship certificate is an original work carried out by R.Praneeth under the guidance of Mr.Charan. This matter embodied in this project is a genuine work, done by me.

**Rikka praneeth**

**ABSTRACT**

This project is a simple user interface which is used to predict the life expectancy of a particular system. This interface used anywhere tp predict the life expectancy of a particular country. This project is based on machine learning which is more accurate in predicting life expectancy rather than many methods that were used in past. This project uses different factors for predicting life expectancy. Factors that are useful for this project are Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. This project simply reduce the burden for many employs as life expectancy is one of the important for countries development evaluation.

**INTRODUCTION**

A typical Regression **Machine Learning** project leverages historical data to predict insights in to the future. This problem statement is aimed at predicting life expectancy rate of a country given various features.

Life expectancy is a statistical measure of the average time a human being is expected to live. Life expectancy depends on various factors: Regional variations, Economic circumstances Differences, Mental illness, Physical illness,Education,Year of the birth

and other demographic factors. This problem statement provides a way to predict average life expectancy of people living in a country when various factors such as year,GDP,education,alcohol intake of people living in a country, expenditure on healthcare system and some specific disease related deaths that happened in the country are given.

**BACKGROUND:**

Life expectancy plays an important role when decisions about the final phase of life need to be made. Good prognostication for example helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more broadly: facilitates Advance Care Planning. Advance Care Planning (ACP) is the process during which patients make decisions about the health care they wish to receive in the future, in case the patient loses the capacity of making decisions or communicating about them. Successful ACP enhances the quality of life and death for palliative patients, by providing timely palliative care and documenting preferences regarding resuscitation and euthanasia, among other things .Accurate prognosis of life expectancy is essential for general practitioners (GPs) to decide when to introduce the topic of ACP to the patient, and it is a key determinant in end-of-life decisions . Increasing the accuracy of prognoses has the potential to benefit patients in various ways by enabling more consistent ACP, earlier and better anticipation on palliative needs, and preventing excessive treatment. This study focuses on automatic life expectancy prediction based on medical records.Although medical records are increasingly available in the form of electronic medical records (EMRs), they remain underutilized for developing clinical decision support systems, and improving health care in general. EMRs are characterized by irregularly-sampled time-series data, missing values, long-term dependencies involving symptoms, diagnoses and interventions, and are prone to documentation errors. Moreover, they contain important information in the form of unstructured, textual data, from which information cannot be extracted straightforwardly. These challenges lead to sub optimal use and even waste of large portions of data, especially when the data is unstructured and noisy. Free texts make up a significant and important part of EMR data, but their ambiguous and noisy character and the and lack of canonical forms for medical concepts and the relations between them make it difficult to ‘mine’ these texts effectively.

**DIFFICULT IN BACKGROUND:**

Accurate prognosis is notoriously difficult; a systematic review investigating the accuracy of clinicians’ estimates of survival of palliative patients shows that there is wide variation in the accuracy of predictions. Although there is a variety of tools available for identifying palliative patients, such as RADPAC , SPICT , and the Surprise Question , virtually none of them are widely used, because using them is time-consuming, and psychological or social factors tend to be marginalized in these tools, although they are important when making end-of-life decisions . In practice, the most important indicators used by GPs when making prognoses tend to be discharge letters from the hospital, increased need for medical care, and decreased social contacts. Identification of patients in need of palliative care depends heavily on the experience of a doctor with palliative patients. Christakis and Lamont investigated the accuracy of doctors in a hospice setting: whenever a new patient was admitted to a participating hospice, a survey with the referring doctor was executed in order to obtain their life expectancy prediction for this patient. Allowing an error margin of 33percent before and after the actual moment of death, the study showed that 20 percent of the life expectancy prognoses were correct. In line with the other studies discussed in, doctors systematically overestimated actual life expectancy – their predictions were too optimistic. Being overoptimistic about life expectancy hinders proper end-of-life care: it may be the root cause of late hospice referral. While experts agree that terminally ill patients should ideally receive 3 months of hospice care, patients in practice usually receive no more than 1 month.

**CHANGES WE NEED:**

Life expectancy is one of the most important factors in end-of-life decision making. Good prognostication for example helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more broadly: facilitates Advance Care Planning. Advance Care Planning improves the quality of the final phase of life by stimulating doctors to explore the preferences for end-of-life care with their patients, and people close to the patients. Physicians, however, tend to overestimate life expectancy, and miss the window of opportunity to initiate Advance Care Planning. **This project tests the potential of using machine learning and natural language processing techniques for predicting life expectancy from electronic medical records.**

**OBJECTIVES:**

The main objective of the project is to predict the life expectancy in more accurate way then the previous methods.

**REQUIREMENTS:**

* **FUNCTIONAL REQUIREMENTS:**

data set,EDA etc.

* **NON-FUNCTIONAL REQUIREMENTS:**

IBM Watson,node-red,processor etc.

**Overall Description:**

The main aim of project is to create a user interface which can be helpful for predicting life expectancy of particular country.Inorder to achieve that one we use IBM cloud platform services. IBM clouds platform services provides us to implement machine learning services in associate with Watson studios by using those services we create project in Watson studios where we can import our dataset as asset and add Auto AI services to project then we select the predicting column after that we can deploy our model. But our project is not yet ready.

Another main step in our project is to create user interface in associate with deployed ML project which we create in Watson studios. To achieve that one, we use node-red service in IBM platform. In node-red we use form node to create UI,function node for input,wbm node to integrate ml service to UI,guard node for displaying predicted value.

**Data Flow DIAGRAM:**



**Sample code :**

|  |
| --- |
| {"id":"265dd308.76498c","type":"tab","label":"Flow 3","disabled":false,"info":""},{"id":"116f2abe.69fe75","type":"wml","z":"265dd308.76498c","name":"rp","connection":"a215b92d.e98fb8","wml-mode":"runPrediction","model":"ea44e0db-09fa-4894-98ee-d6be1a6c4ef0","deployment":"e0de2dc2-523f-4938-ac9e-a8b15bfa9ea3","modelhidden":"ea44e0db-09fa-4894-98ee-d6be1a6c4ef0","deploymenthidden":"","x":430,"y":260,"wires":[["909a4939.ea7dd8","2305e492.8a128c","aab9adc5.f5087","4e78cfa1.87e6b"]]},{"id":"d90bd0d9.0093a","type":"ui\_form","z":"265dd308.76498c","name":"","label":"","group":"7cb91b07.f8af64","order":1,"width":0,"height":0,"options":[{"label":"country","value":"co","type":"text","required":true,"rows":null},{"label":"year","value":"ye","type":"number","required":true,"rows":null},{"label":"status","value":"st","type":"text","required":true,"rows":null},{"label":"adult mortality","value":"am","type":"number","required":true,"rows":null},{"label":"infant deaths","value":"ind","type":"number","required":true,"rows":null},{"label":"Alcohol","value":"al","type":"number","required":true,"rows":null},{"label":"percentage expenditure","value":"pe","type":"number","required":true,"rows":null},{"label":"Hepatitis B","value":"he","type":"number","required":true,"rows":null},{"label":"Measles","value":"me","type":"number","required":true,"rows":null},{"label":"BMI","value":"bm","type":"number","required":true,"rows":null},{"label":"under five deaths","value":"un","type":"number","required":true,"rows":null},{"label":"Polio","value":"pol","type":"number","required":true,"rows":null},{"label":"Total expenditure","value":"te","type":"number","required":true,"rows":null},{"label":"Diphtheria","value":"di","type":"number","required":true,"rows":null},{"label":"HIV/AIDS","value":"ha","type":"number","required":true,"rows":null},{"label":"GDP","value":"gd","type":"number","required":true,"rows":null},{"label":"Population","value":"po","type":"number","required":true,"rows":null},{"label":"thinness 1-19 years","value":"th","type":"number","required":true,"rows":null},{"label":"thinness 5-9 years","value":"thi","type":"number","required":true,"rows":null},{"label":"Income composition of resources","value":"inc","type":"number","required":true,"rows":null},{"label":"Schooling","value":"sc","type":"number","required":true,"rows":null}],"formValue":{"co":"","ye":"","st":"","am":"","ind":"","al":"","pe":"","he":"","me":"","bm":"","un":"","pol":"","te":"","di":"","ha":"","gd":"","po":"","th":"","thi":"","inc":"","sc":""},"payload":"","submit":"submit","cancel":"cancel","topic":"","x":150,"y":160,"wires":[["54ab4da7.c60d54"]]},{"id":"909a4939.ea7dd8","type":"debug","z":"265dd308.76498c","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":690,"y":180,"wires":[]},{"id":"54ab4da7.c60d54","type":"function","z":"265dd308.76498c","name":"","func":"msg.payload = {values:[[\nmsg.payload.co,msg.payload.ye,msg.payload.st,msg.payload.am,msg.payload.ind,msg.payload.al,msg.payload.pe,msg.payload.he,msg.payload.me,msg.payload.bm,msg.payload.un,msg.payload.pol,msg.payload.te,msg.payload.di,msg.payload.ha,msg.payload.gd,msg.payload.po,msg.payload.th,msg.payload.thi,msg.payload.inc,msg.payload.sc],[\nmsg.payload.co,msg.payload.ye,msg.payload.st,msg.payload.am,msg.payload.ind,msg.payload.al,msg.payload.pe,msg.payload.he,msg.payload.me,msg.payload.bm,msg.payload.un,msg.payload.pol,msg.payload.te,msg.payload.di,msg.payload.ha,msg.payload.gd,msg.payload.po,msg.payload.th,msg.payload.thi,msg.payload.inc,msg.payload.sc]]}\nreturn msg;","outputs":1,"noerr":0,"x":290,"y":260,"wires":[["116f2abe.69fe75"]]},{"id":"5c1fc0e7.b4be8","type":"ui\_text","z":"265dd308.76498c","group":"7cb91b07.f8af64","order":2,"width":0,"height":0,"name":"","label":"life expectancy","format":"{{msg.payload}}","layout":"row-spread","x":720,"y":100,"wires":[]},{"id":"2305e492.8a128c","type":"function","z":"265dd308.76498c","name":"global","func":"global.set(\"pred\",msg.payload.predictions[0].values[1])\nreturn msg;","outputs":1,"noerr":0,"x":490,"y":360,"wires":[[]]},{"id":"aab9adc5.f5087","type":"function","z":"265dd308.76498c","name":"pred","func":"var pred=global.get(\"pred\")\nreturn {payload:msg.payload.predictions[0].values[1]}","outputs":1,"noerr":0,"x":632,"y":385,"wires":[["924507da.a55588"]]},{"id":"924507da.a55588","type":"ui\_gauge","z":"265dd308.76498c","name":"","group":"7cb91b07.f8af64","order":2,"width":0,"height":0,"gtype":"gage","title":"life prediction","label":"percentage","format":"{{value}}","min":0,"max":"100","colors":["#00b500","#e6e600","#ca3838"],"seg1":"","seg2":"","x":810,"y":340,"wires":[]},{"id":"4e78cfa1.87e6b","type":"function","z":"265dd308.76498c","name":"pred","func":"var pred=global.get(\"pred\")\nreturn {payload:msg.payload.predictions[0].values[1]}","outputs":1,"noerr":0,"x":530,"y":120,"wires":[["5c1fc0e7.b4be8"]]},{"id":"a215b92d.e98fb8","type":"wml-config","z":"","host":"https://eu-gb.ml.cloud.ibm.com","instanceid":"75f8411c-ec24-4915-8a1d-6f2559e48f12","name":"wdp-writer"},{"id":"7cb91b07.f8af64","type":"ui\_group","z":"","name":"rp","tab":"62a9f009.03709","order":3,"disp":true,"width":"6","collapse":false},{"id":"62a9f009.03709","type":"ui\_tab","z":"","name":"chef","icon":"dashboard","disabled":false,"hidden":false}] |

**Screenshots:**



**Fig:1 node-red flow**

 Fig2:user interface



**Fig 3: user interface**



Fig 4: output

**Conclusion:**

Finally, we conclude that this user interface is very useful in predicting life expectancy with more accurate than other old methods used in the past.