



RV Educational Institutions[®]
RV College of Engineering[®]

Autonomous
Institution Affiliated
to Visvesvaraya
Technological
University, Belagavi

Approved by AICTE,
New Delhi, Accredited
By NAAC, Bengaluru
And NBA, New Delhi

College	RV College of Engineering, Bengaluru		
Department	Department of Computer Science and Engineering		
Course: Course code:	18CS6C5 - BIG DATA ANALYTICS USING DISTRIBUTED PLATFORMS		
USN Nos.	1RV18CS028 1RV18CS163 1RV18TE057	Student Names	Aravind A Shubh Shukla V Rahul
Project Title	A comparative study of Hadoop and HPCC frameworks		
Guide	Dr. Shobha G, Professor, R.V College of Engineering		

SYNOPSIS

HPCC systems are powerful, open-sourced big data technologies in the market today. They are licensed under Apache 2.0 license. They make use of commodity hardware and local storage interconnected through IP networks which allows parallel data processing and/or querying across the architectures. HPCC Systems essentially use a higher-level programming language called enterprise control language (ECL), which is based on C++. Being a declarative language ECL is easy to work with for people coming from other areas other than traditional programming.

We being newcomers to the field of big data, we wanted to explore machine learning modules on HPCC Systems to better understand the architecture and get the flow going with ECL. We are using Classification decision tree machine learning algorithm with HPCC Systems to classify whether the person can be diagnosed with COVID-19. We are going to make use of covid dataset from <https://data.gov.il/dataset/covid-19> to make our predictions using the following variables.

- Age over 60 – Age_60
- Sex - Male (Male=1, Female=0)
- Cough
- Shortness of breath
- Fever
- Sore throat
- Headache
- Contact with a confirmed individual

Outcome: The probability of being diagnosed with a COVID-19 infection.