**Project Proposal**

1. **Title of the Project: Traffic Accident Analysis and Recommendations**
2. **Brief on the project:**

Traffic accidents cause a staggering loss to life and property of citizens. One of the primary objectives of the National Highway Traffic Safety Administration (NHTSA) is to reduce the toll that motor vehicle accidents cause on society. NHTSA uses data from many sources, including the Fatality Analysis Reporting System (FARS) which began operation in 1975. Providing data

about fatal crashes involving all types of vehicles, the FARS is used to identify highway safety

problem areas, provide a basis for regulatory and consumer information initiatives, and form the basis for cost and benefit analyses of highway safety initiatives.

FARS is a census of fatal motor vehicle crashes with a set of data files documenting all qualifying fatalities that occurred within the United States since 1975. To qualify as a FARS case, the crash had to involve a motor vehicle traveling on a trafficway customarily open to the public, and must have resulted in the death of a motorist or a non-motorist within 30 days of the crash.

FARS data are obtained from various States’ documents, such as:

*Police Accident Reports (PAR), Death Certificates, State Vehicle Registration Files, Coroner/Medical Examiner Reports, State Driver Licensing Files, State Highway Department Data, Emergency Medical Service Reports, Vital Statistics and other State Records*

1. **Deliverables of the project:** The data that has been made available in FARS for 2015 contains 20 files, key ones being:
2. Accident (52 features, 32k records)
3. Vehicle (102 features, 49k records)
4. Person (68 features, 81k records)
5. Distractions (4 features, 49k records)
6. Impairments (4 features, 49k records)
7. Factors (4 features, 49k records)
8. Vision (4 features, 49k records)

The first task would be to clean and merge the data into a coherent dataset. Since the data has been coded into FARS by trained enumerators from their primary sources, it is expected that the data will be of good quality. Further the relevant features would have to be identified and shortlisted. Patterns or clusters will be identified from this data. The deliverables from this project will be:

1. Analysis presented using a visualization tool
2. A model that can be used to predict the number of fatalities in an accident

Key questions that the project will answer:

1. Is there a pattern to fatal accidents in terms of time of day, day of week, month, holidays etc
2. Ranking regions/states on the basis of accidents and fatalities
3. Impact of weather/visibility on accidents and fatalities
4. Impact of distractions like mobile devices, eating/drinking, smoking, music/video
5. Impact of physical or mental impairments like handicap, fatigue, depression
6. Impact of emergency services response time, time taken to reach hospital etc
7. Impact of drugs and alcohol
8. Impact of usage/non-usage of safety equipment like restraints, air bags etc
9. **Resources**

* **Data set source:** <ftp://ftp.nhtsa.dot.gov/fars/2015/>
* **Soft ware**: R Studio, Python, Tableau (Or any other visualization tools), Shiney Or HTML for website creation
* **References**:<https://www.aaai.org/Papers/Symposia/Spring/2009/SS-09-04/SS09-04-002.pdf> <http://www.sciencedirect.com/science/article/pii/S1877042814001074> <http://www.vegvesen.no/_attachment/336339/binary/585485>

<http://www.softcomputing.net/informatica2.pdf>

1. **Team Members:** Johnson Varghese, [johnnyv4@rediffmail.com](mailto:johnnyv4@rediffmail.com), 9821285034

Parag Mehta, [parag.mehta83@yahoo.com](mailto:parag.mehta83@yahoo.com), 8452045528

Ranjeet Singh Rawat, [ranjeet143@gmail.com](mailto:ranjeet143@gmail.com), 9665040690

Bhavesh Vani, [bhavesh\_vani@yahoo.com](mailto:bhavesh_vani@yahoo.com), 8691963652

1. **Milestones with timeline**

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestones** | **Start date** | **Finish date** | **Grade Grading** |
| 1. **Define a problem** |  | **15-Sep** | **5%** |
| 1. **Get the Data** |  | **15-Sep** | **5%** |
| 1. **Explore and pre-process data** | **16-Sep** | **10-Oct** | **10%** |
| 1. **Create Features** | **11-Oct** | **15-Nov** | **5%** |
| 1. **Create Model** | **11-Oct** | **15-Nov** | **15%** |
| 1. **Deploy & consume model** | **16-Nov** | **30-Nov** | **10%** |
| 1. **Report Writing** | **1-Dec** | **15-Dec** | **10%** |
| 1. **Project submission** |  | **15-Dec** | **15%** |
| 1. **Final presentation with project/ product demonstration** |  | **17-Dec** | **20%** |
| 1. **Blog publishing** | **1-Dec** | **18-Dec** | **5%** |

1. **Technical architecture**

