****

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

**Project Review – I**

1. **Project title and project member details**

|  |  |  |  |
| --- | --- | --- | --- |
| **Batch Size** | **Reg No** | **Name** | **Title of the Project** |
|
| 4 | 15BCE1326 | OSHO AGYEYA | DEEP LEARNING ON GRAPH DATABASE USING NEO4J AND KERAS |
| 15BCE1252 | TUSHAR PAHUJA |
| 15BCE1257 | HARGUR BEDI |
| 15BCE1136 | PUSHPIT BAGGA |

1. **Abstract**

**Title:** Deep learning on graph database using Neo4j and Keras

**Neo4j** is a graph database management system developed by Neo4j, Inc. Described by its developers as an ACID-compliant transactional database with native graph storage and processing, Neo4j is the most popular graph database according to DB-Engines ranking, and the 22nd most popular database overall.

**Keras** is an open source neural network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit or Theano. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible.

The project is aimed to connect a Neo4j graph database to Keras. The main objective of the project is to create a neural network for a review prediction task. The database is created in Neo4j and it is injected into Keras learing model to make predictions regarding reviews of products.

1. **Concept modules of the project**
2. **List of Modules**
3. Creation of graph database manually in Neo4j
4. Exploratory data analysis
5. Establishing connection between Neo4j and Keras
6. Create a machine learning model
7. Training session of the model
8. Hyperparameter tuning after observing results of training session
9. Finalizing the model
10. Making final predictions
11. **Concepts Applied in the Project**

**Graph database:** Very simply, a graph database is a database designed to treat the relationships between data as equally important to the data itself. It is intended to hold data without constricting it to a pre-defined model. Instead, the data is stored like we first draw it out – showing how each individual entity connects with or is related to others. If you do decide to move your data from a relational to a graph database, the steps to transition your applications to use Neo4j are actually quite simple. You can connect to Neo4j with a driver or connector library designed for your stack or programing language, just as you can with other databases. Thanks to Neo4j and its community, there are Neo4j drivers that mimic existing database driver idioms and approaches for nearly any popular programing language.

**Deep learning:** Deep learning (also known as deep structured learning or hierarchical learning) is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised.

Deep learning architectures such as deep neural networks, deep belief networks and recurrent neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design and board game programs, where they have produced results comparable to and in some cases superior to human experts.

1. **Platform & Language to be used for Implementation**

**Neo4j:** It is a [graph database](https://en.wikipedia.org/wiki/Graph_database) management system developed by Neo4j, Inc. Described by its developers as an [ACID](https://en.wikipedia.org/wiki/ACID)-compliant transactional database with native graph storage and processing, Neo4j is the most popular graph database according to [DB-Engines ranking](https://en.wikipedia.org/wiki/DB-Engines_ranking), and the 22nd most popular database overall.

**Keras:** It is an [open source](https://en.wikipedia.org/wiki/Open-source_software) [neural network](https://en.wikipedia.org/wiki/Artificial_neural_network) library written in [Python](https://en.wikipedia.org/wiki/Python_(programming_language)). It is capable of running on top of [TensorFlow](https://en.wikipedia.org/wiki/TensorFlow), [Microsoft Cognitive Toolkit](https://en.wikipedia.org/wiki/Microsoft_Cognitive_Toolkit) or [Theano](https://en.wikipedia.org/wiki/Theano_(software)). Designed to enable fast experimentation with [deep neural networks](https://en.wikipedia.org/wiki/Deep_learning), it focuses on being user-friendly, modular, and extensible.

**Python:** It is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. In July 2018, Van Rossum stepped down as the leader in the language community after 30 years. Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

**Cypher:** It is a declarative graph query language that allows for expressive and efficient querying and updating of a property graph. Cypher is a relatively simple but still very powerful language. Very complicated database queries can easily be expressed through Cypher. This allows users to focus on their domain instead of getting lost in database access.