The present work of thesis mainly presents some innovative data-driven applications to deal with the emerging data challenges in power system.

With the overwhelming trend of digitalization in power system, critical challenges and opportunities are emerging in the modern industrial systems. In the first chapter, the big data analytics are introduced with corresponding applications in smart grids. With huge amount of data from electricity network, meteorological information system, geographical information system etc., many benefits can be brought to the existing network and import the customer service as well as the social welfare in the era of big data.

In distribution network, there are a large number of feeders spread in a wide area with a diversity of structural features. Chapter 2 discusses the possibility to group thousands of feeders with a data-driven method and try to identify the representative feeders for a better evaluation of the performance with regard to the number of outages. Two clustering algorithms are applied on the mixed data with the similar analysis results.

In Chapter 3, the data-driven solutions to the problems relating to the outage predictions in medium-voltage distribution network is presented. Annual and monthly number of outages are predicted based on improved grey theory. Further more, the daily outages prediction is discussed as a binary classification problem with support vector machine.

Facing the challenges of limited channel bandwidth in advanced metering infrastructures, the consumption records with partial missing information are utilized in Chapter 4 to figure out the abnormal users in power system. By extracting the importance index through regression models, anomalies from the energy consumption patterns are successfully uncovered.

Finally, the Monte Carlo simulation is introduced to evaluate the high-impact low-probability events in the local distribution network in Chapter 5. Since there are considerable increase of outages in the summer, heat waves are regarded as a severe threat to the power system security. A detailed analysis between the temperature and number of outages is firstly discussed as the basis of probabilistic simulation. The savings of expenditures due to a decrease of repetitive faults after investment could be estimated via calculation from thousands of scenarios.