# PENGUJIAN SISTEM

Pada bab ini, akan dibahas pengujian terhadap sistem NEAT dan *backpropagation* yang telah dibuat.

1. **Sistem Pengujian**

Waktu pengujian pada bab ini sangat ditentukan oleh spesifikasi yang digunakan. Semua pengujian yang dilakukan pada bab ini dilakukan menggunakan *notebook asus A412DA* dengan spesifikasi sebagai berikut:

Tabel 5.1 Spesifikasi sitem pengujian

|  |  |
| --- | --- |
| *Processor* | *AMD Ryzen 5 3500u* |
| *RAM* | *8GB DDR4* |
| *Operating System* | *Windows 10* |

1. **Pengujian NEAT**

Pengujian NEAT pada penelitian ini akan dibagi kedalam 3 tahap berdasarkan *feature* yang digunakan. Pembagian ini bertujuan untuk melihat pengaruh dari *feature* yang digunakan terhadap akurasi dari NEAT. Pada setiap tahap, akan dicoba berbagai konfigurasi untuk mendapatkan hasil yang optimal. Pengujian yang dilakukan meliputi proses *training*, yang bertujuan untuk mencari individu terbaik, dan proses *testing*, untuk mengukur akurasi dari individu yang dihasilkan oleh proses *training*.

Data yang digunakan berasal dari situs *whosocred.*com, yang merupakan situs penyedia data pertandingan sepak bola. Pada proses *training*, data berasal dari pertandingan liga inggris pada musim 2014/2015, 2015/2016, dan 2016/2017. Sedangkan untuk proses *testing*, data yang digunakan berasal dari pertandingan liga inggris pada musim 2017/2018. Pada setiap musim, terdapat 380 pertandingan.

### Tahap 1

Pada tahap ini, *feature* yang digunakan adalah *rating* pemain dari kedua *team*, sehingga ada 22 *feature* yang terdiri dari 11 pemain dari masing-masing *team*.

#### Tahap 1 Pengujian 1

A screenshot of text

Description automatically generated

Gambar 5.1 Konfigurasi yang digunakan pada Tahap 1 Pengujian 1

Seperti yang sudah dijelaskan pada Bab 3 dan 4, tahap pertama pada *training* NEAT adalah pengaturan konfigurasi. Konfigurasi yang dilakukan pada pengujian ini dapat dilihat pada Gambar 5.1. Setelah konfigurasi, proses selanjutnya adalah pemilihan *fitness function*. Untuk pengujian pertama ini, *fitness function* yang digunakan adalah *fitness function* pertama, yang dapat dilihat pada Segmen Program 4.4.

A screenshot of a cell phone

Description automatically generated

Gambar 5.2 Hasil training dari tahap 1 pengujian 1

Setelah dijalankan sebanyak 5000 generasi dengan waktu rata-rata 6.8 detik per generasi, total waktu yang dibutuhkan NEAT untuk menjalankan proses *training* dengan tahap 1 pengujian 1 adalah 571.75 menit. Hasil pengujian dapat dilihat pada Gambar 5.2. Dengan konfigurasi dan *fitness function* yang digunakan, individu terbaik yang dihasilkan NEAT memiliki nilai fitness sebesar 555.5, dengan rata-rata *fitness* pada populasi sebesar 259.39311. Grafik *fitness* terbaik selama pengujian dan rata-rata *fitness* dalam populasi dapat dilihat pada Gambar 5.3.

A close up of a piece of paper

Description automatically generated

Gambar 5.3 Grafik nilai fitness terbaik dan nilai fitness rata-rata dalam populasi selama proses training pada Tahap 1 Pengujian 1

Individu terbaik yang dihasilkan oleh NEAT dapat dilihat pada Gambar 5.4. Individu ini memiliki kompleksitas jaringan berupa 22 *input* *nodes*, 9 *hidden nodes*, dan 2 *output nodes* dengan 23 koneksi aktif. Dimana *nodes* -1 sampai -22 merupakan *input* *nodes*, 0 dan 1 merupakan *output nodes*, dan sisanya merupakan *hidden nodes*. Garis merah dan hijau menandakan koneksi antar *nodes*, yang berarti *weight* positif untuk garis hijau dan negatif untuk garis merah. Sedangkan garis putus-putus menandakan koneksi yang tidak aktif.

A close up of a map

Description automatically generated

Gambar 5.4 Network terbaik yang dihasilkan NEAT pada Tahap 1 Pengujain 1

Setelah *training* selesai, proses selanjutnya adalah *testing* yang bertujuan untuk mengukur tingkat akurasi dari *network* terbaik yang dihasilkan NEAT. Namun setelah dilakukannya proses *testing*, tingkat akurasi dari *network* terbaik ini masih sangat buruk. *Network* hanya mampu memprediksi hasil pertandingan dengan benar sebanyak 107 pertandingan, dimana 32 diantaranya dapat diprediksi dengan skor yang tepat. Ringkasan proses *training* dan proses *testing* dapat dilihat pada Tabel 5.2.

Tabel 5.2 Rangkuman proses training dan testing Tahap 1 Pengujian 1

|  |  |
| --- | --- |
| Generasi | 5000 |
| Rata-rata waktu per generasi | 6.861 detik |
| Total waktu | 571.75 menit |
| *Fitness Function* | *Fitness Function 1* (Segmen Program 4.4) |
| Rata-rata nilai *fitness* dalam populasi | 259.39311 |
| Nilai *fitness* terbaik dalam populasi | 555.5 |
| Kompleksitas *network* pada individu dengan nilai *fitness* terbaik | 33 *nodes* dan 23 koneksi aktif |
| Skor pertandingan benar | 32 (8%) |
| Hasil pertandingan benar | 107 (28%) |

#### 5.2.1.2 Tahap 1 Pengujian 2

Karena hasil Tahap 1 Pengujian 1 kurang memuaskan, pada pengujian ini *fitness function* yang digunakan diganti menjadi *fitness function* *2*, yang dapat dilihat pada Segmen Program 4.2.

Sedangkan untuk konfigurasi yang digunakan, masih sama dengan konfigurasi yang digunakan pada Tahap 1 Pengujian 1.

A screenshot of text

Description automatically generated

Gambar 5.5 Konfigurasi yang digunakan pada Tahap 1 Pengujian 2

Proses *training* pada pengujian ini juga dijalankan sebanyak 5000 generasi. Namun, pengujian ini memakan waktu yang lebih lama dibanding Tahap 1 Pengujian 1. Rata-rata waktu per generasi yang dibutuhkan pada pengujian ini adalah 8.124 detik, sehingga total waktu yang dibutuhkan untuk menyelesaikan proses *training* adalah 677 menit. Hasil *training* dapat dilihat pada Gambar 5.6

A screenshot of a cell phone

Description automatically generated

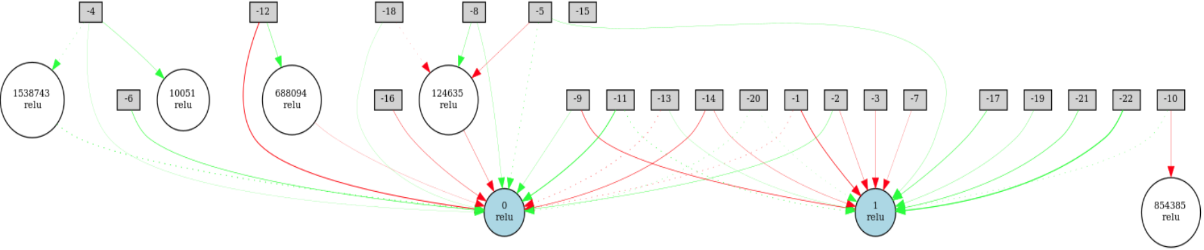
Gambar 5.6 Hasil training dari Tahap 1 Pengujian 2

Setelah proses *training* selesai, individu yang keluar sebagai pemenang memiliki nilai *fitness* sebesar -272.275. Perlu diingat bahwa *fitness* *function­* yang digunakan pada penelitain ini berbeda dengan Tahap 1 Pengujian 1, sehingga nilai *fitness* pada penelitian ini tidak bisa dibandingkan secara langsung dengan nilai *fitness* terbaik yang ada pada Tahap 1 Pengujian 1. Grafik *fitness* terbaik selama pengujian dan rata-rata *fitness* dalam populasi dapat dilihat pada Gambar 5.7.

A screenshot of a cell phone

Description automatically generated

Gambar 5.7 Grafik nilai fitness terbaik dan nilai fitness rata-rata dalam populasi selama proses training pada Tahap 1 Pengujian 2



Gambar 5.8 Network terbaik yang dihasilkan NEAT pada Tahap 1 Pengujain 2

Setelah dilakukan proses testing, *network* yang dihasilkan NEAT ternyata memiliki tingkat akurasi yang jauh lebih tinggi jika dibandingkan dengan Tahap 1 Pengujian 1. *Network* pada pengujian ini mampu memprediksi hasil pertandingan sebanyak 273 dari 380 pertandingan, dimana 85 diantaranya dapat diprediksi dengan skor yang benar. Hal ini membuktikan bahwa *fitness function 2* mampu menilai sebuah *network* lebih baik daripada *fitness function 1* karena konfigurasi yang digunakan pada pengujian ini sama dengan konfigurasi yang digunakan pada Tahap 1 Pengujian 1. Rangkuman Tahap 1 Pengujian 2 dapat dilihat pada Tabel 5.3.

Tabel 5.3 Rangkuman proses training dan testing Tahap 1 Pengujian 2

|  |  |
| --- | --- |
| Generasi | 5000 |
| Rata-rata waktu per generasi | 8.124 detik |
| Total waktu | 677 menit |
| *Fitness Function* | *Fitness Function 2* (Segmen Program 4.5) |
| Rata-rata nilai *fitness* dalam populasi | -7509.48 |
| Nilai *fitness* terbaik dalam populasi | -272.275 |
| Kompleksitas *network* pada individu dengan nilai *fitness* terbaik | 29 *nodes* dan 29 koneksi aktif |
| Skor pertandingan benar | 85 (22%) |
| Hasil pertandingan benar | 273 (71%) |

#### Tahap 1 Pengujian 3

Berdasarkan hasil dari Tahap 1 Pengujian 2, *fitness* *function* yang akan digunakan pada pengujian ini adalah *fitness function 2* karena sudah terbukti lebih baik daripada *fitness function 1*.

Konfigurasi pada pengujian ini sedikit berbeda dengan 2 pengujian sebelumnya. Pada pengujian ini, conn\_delete\_prop dan node\_delete\_prob yang mengatur kemungkinan untuk hilangnya koneksi dan *nodes* dalam proses *mutation* diubah menjadi 0 dari 0.5 seperti pada 2 pengujian sebelumnya. Dengan dilakukannya perubahan konfigurasi ini, diharapkan individu dapat berkembang lebih cepat sehingga mampu menghasilkan akurasi yang lebih baik.

A screenshot of text

Description automatically generated

Gambar 5.9 Konfigurasi yang digunakan pada Tahap 1 Pengujian 3

A screenshot of a cell phone

Description automatically generated

Gambar 5.10 Hasil training dari Tahap 1 Pengujian 3

Efek perubahan konfigurasi dapat dilihat pada hasil *training* pengujian ini pada Gambar 5.10. Dengan mengatur kemungkinan hilangnya koneksi dan *nodes* menjadi 0, koneksi yang dimiliki *network* terbaik pada pengujian ini menjadi jauh lebih banyak dan nilai *fitness* terbaik juga lebih tinggi dari 2 pengujian sebelumnya.

Grafik *fitness* terbaik selama pengujian dan rata-rata *fitness* dalam populasi dapat dilihat pada Gambar 5.3.

A screenshot of a cell phone

Description automatically generated

Gambar 5.11 Grafik nilai fitness terbaik dan nilai fitness rata-rata dalam populasi selama proses training pada Tahap 1 Pengujian 1

Tidak hanya efek positif, perubahan konfigurasi ini juga disertai efek negatif. Pada 2 pengujian sebelumnya, 1 generasi dapat diselesaikan dalam waktu kurang dari 10 detik. Tetapi pada pengujian ini, diibutuhkan rata-rata 18.8 detik untuk menyelesaikan 1 generasi. Sehingga total waktu yang dibutuhkan untuk menyelesaikan 5000 generasi adalah sekitar 1566 menit, jauh lebih lama dari 2 pengujian sebelumnya.

Efek positif yang ditemukan pada proses *training* juga berpengaruh pada proses *testing*. Tingkat akurasi dari *network* yang dihasilkan pada pengujian ini mampu melebihi tingkat akurasi pada 2 pengujian sebelumnya. *Network* pada pengujian ini mampu memprediksi 300 dari 380 pertandingan dengan benar, dimana 125 diantaranya dapat diprediksi dengan skor yang benar. Rangkuman proses *training* dan *testing* dapat dilihat pada Tabel 5.4.

Tabel 5.4 Rangkuman proses training dan testing pada Tahap 1 Pengujian 3

|  |  |
| --- | --- |
| Generasi | 5000 |
| Rata-rata waktu per generasi | 18.803 detik |
| Total waktu | 1566 menit |
| *Fitness Function* | *Fitness Function 2* (Segmen Program 4.5) |
| Rata-rata nilai *fitness* dalam populasi | -12051.081 |
| Nilai *fitness* terbaik dalam populasi | 75.796 |
| Kompleksitas *network* pada individu dengan nilai *fitness* terbaik | 36 *nodes* dan 70 koneksi aktif |
| Skor pertandingan benar | 125 (30%) |
| Hasil pertandingan benar | 300 (78%) |

### 5.2.2 Tahap 2

Pada tahap ini, *feature* yang digunakan adalah *rating* pemain dari kedua *team*, dan *rating* kedua team itu sendiri. Sehingga ada total 24 *feature* yang terdiri dari 11 pemain dari masing-masing *team* dan *rating* dari kedua team.

#### Tahap 2 Pengujian 1

Pada pengujian ini, konfigurasi dan *fitness function* yang digunakan sama seperti Tahap 1 Pengujian 3, hanya saja num\_inputs berubah menjadi 24 dari yang sebelumnya 22 karena adanya penambahan 2 *feature* baru.

A screenshot of text

Description automatically generated

Gambar 5.12 Konfigurasi yang digunakan pada Tahap 2 Pengujian 1

A screenshot of a cell phone

Description automatically generated

Gambar 5.13 Hasil training dari Tahap 2 Pengujian 1

Sama seperti proses *training* pada pengujian sebelumnya, *training* pada pengujian ini juga dijalankan selama 5000 generasi. Penambahan *feature* baru pada pengujian ini ternyata tidak memiliki banyak efek positif pada proses *training*. Dapat dilihat pada Gambar 5.13., nilai *fitness* terbaik yang dihasilkan ternyata lebih kecil daripada nilai *fitness* yang dihasilkan oleh Tahap 1 Pengujian 3. Grafik *fitness* terbaik dan rata-rata *fitness* dalam populasi dapat dilihat pada Gambar 5.14.

A screenshot of a cell phone

Description automatically generated

Gambar 5.14 Grafik nilai fitness terbaik dan nilai fitness rata-rata dalam populasi selama proses training pada Tahap 2 Pengujian 1

Efek positif penambahan 2 *feature* pada pengujian ini baru terlihat pada setelah proses *testing* selesai dijalankan. Tingkat akurasi dari *network* yang dihasilkan pengujian ini sedikit lebih tinggi daripada Tahap 1 Pengujian 3, yang merupakan *network* dengan tingkat akurasi sejauh ini. *Network* yang dihasilkan pada pengujian mampu memprediksi pertandingan dengan benar sebanyak 302 dari 380 pertandingan, dimana 136 diantaranya dapat diprediksi dengan skor yang tepat. *Network* pada pengujian ini juga memiliki tingkat kompleksitas sedikit lebih rendah dari Tahap 1 Pengujian 3. Rangkuman pengujian ini dapat dilihat pada Tabel 5.5.

Tabel 5.5 Rangkuman proses training dan testing Tahap 2 Pengujian 1

|  |  |
| --- | --- |
| Generasi | 5000 |
| Rata-rata waktu per generasi | 16.769 detik |
| Total waktu | 1397 menit |
| *Fitness Function* | *Fitness Function 2* (Segmen Program 4.5) |
| Rata-rata nilai *fitness* dalam populasi | -11662.11922 |
| Nilai *fitness* terbaik dalam populasi | 50.41736 |
| Kompleksitas *network* pada individu dengan nilai *fitness* terbaik | 33 *nodes* dan 61 koneksi aktif |
| Skor pertandingan benar | 136 (35%) |
| Hasil pertandingan benar | 302 (79%) |

#### Tahap 2 Pengujian 2

A screenshot of text

Description automatically generated

Gambar 5.15 Konfigurasi yang digunakan pada Tahap 2 Pengujian 2

Pada 2 pengujian sebelumnya, menghilangkan kemampuan individu untuk kehilangan *nodes* dan koneksi pada saat terjadinya *mutation* terbukti memberikan efek positif pada tingkat akurasi yang dihasilkan *network*. Pada pengujian ini, node\_add\_prob dan conn\_add\_prop yang mengatur kemungkinan terjadinya penambahan koneksi *nodes* pada proses *mutation* ditingkatkan dari 0.5 menjadi 0.8, dengan harapan *network* dapat berkembang lebih cepat lagi.

A screenshot of a cell phone

Description automatically generated

Gambar 5.16 Hasil training dari Tahap 2 Pengujian 2

A screenshot of a cell phone

Description automatically generated

Gambar 5.17 Grafik nilai fitness terbaik dan nilai fitness rata-rata dalam populasi selama proses training pada Tahap 2 Pengujian 2

Dan ternyata benar, dapat dilihat pada Gambar 5.16., peningkatan kemungkinan terjadinya penambahan koneksi dan *nodes* saat *mutation* memberikan hasil *training* yang lebih baik. Nilai *fitness* yang dihasilkan pada pengujian ini mampu mengungguli nilai *fitness* yang dihasilkan Tahap 2 Pengujian 1 dan Tahap 1 Pengujian 3, yang merupakan 2 *network* terbaik dalam hal nilai *fitness*. Grafik nilai *fitness* terbaik dan rata-rata *fitness* dalam populasi selama proses *training* dapat dilihat pada Gambar 5.17.

Tidak hanya pada proses *training*, efek positif dari konfigurasi yang digunakan juga dapat lihat pada hasil proses *testing*. *Network* yang dihasilkan pengujian ini kembali mampu menghasilkan tingkat akurasi yang lebih tinggi dari pengujian-pengujian sebelumnya. *Network* yang dihasilkan pada pengujian ini mampu memprediksi hasil pertandingan dengan benar sebanyak 308 dari 380 pertandingan, dimana 136 diantarnya dapat diprediksi dengan skor yang tepat.

Tentu saja dengan meningkatkan kemungkinan terjadinya penambahan koneksi dan *nodes* pada saat *mutation* membuat kompleksitas *network* yang dihasilkan menjadi lebih tinggi. Rangkuman proses *training* dan *testing* pada pengujian ini dapat dilihat pada Tabel 5.6

Tabel 5.6 Rangkuman proses training dan testing dari Tahap 2 Pengujian 2

|  |  |
| --- | --- |
| Generasi | 5000 |
| Rata-rata waktu per generasi | 18.802 detik |
| Total waktu | 1566 menit |
| *Fitness Function* | *Fitness Function 2* (Segmen Program 4.5) |
| Rata-rata nilai *fitness* dalam populasi | -11531.72723 |
| Nilai *fitness* terbaik dalam populasi | 116.81931 |
| Kompleksitas *network* pada individu dengan nilai *fitness* terbaik | 44 *nodes* dan 71 koneksi aktif |
| Skor pertandingan benar | 136 (35%) |
| Hasil pertandingan benar | 302(80%) |

### Tahap 3

Pada tahap ini, *feature* yang digunakan kembali ditambah. Selain *rating* dari masing-masing *team* dan pemain, posisi pemain juga digunakan sebagai *feature*. Terdapat 16 posisi pemain pada *dataset* yang digunakan, dimana masing-masing pemain memiliki 1 posisi. Karena posisi pemain tidak berbentuk angka, harus dilakukan *encoding* terlebih dulu sebelum posisi dapat digunakan. Banyaknya *feature* yang digunakan pada tahap ini bergantung pada jenis *encoding* yang digunakan.

#### Tahap 3 Pengujian 1

Pada pengujian pertama dalam tahap 3, jenis *encoding* yang digunakan adalah label *encoding*. Label *encoding* merubah posisi pemain menjadi sebuah nilai tunggal. Dengan menggunakan *encoding* jenis ini, jumlah *feature* yang digunakan adalah 46, yang terdiri dari 22 *rating* pemain, 22 posisi yang telah di*encoding*, dan 2 *rating* dari masing-masing *team*.

A screenshot of text

Description automatically generated

Gambar 5.18 Konfigurasi yang digunakan pada Tahap 3 Pengujian 1

Konfigurasi yang digunakan pada pengujian ini masih sama seperti yang digunakan pada Tahap 2 Pengujian 3, yang merupakan pengujian dengan tingkat akurasi terbaik sejauh ini. Hanya saja, num\_inputs berubah menjadi 46 dari yang sebelumnya 24. Konfigurasi pada pengujian ini dapat dilihat pada Gambar 5.18.

A screenshot of a cell phone

Description automatically generated

Gambar 5.19 Hasil training dari Tahap 3 Pengujian 1

Penambahan *feature* pada pengujian ini membawa banyak efek negatif. Seperti yang dapat dilihat pada Gambar 5.19., waktu yang dibutuhkan untuk menyelesaikan 1 generasi melebihi 20 detik, yang mana tidak pernah terjadi pada penelitian sebelumnya. Untuk menyelesaikan proses *training* sebanyak 5000 generasi, dibutuhkan kurang lebih 2234.08333 menit.

Efek negatif juga dirasakan pada nilai *fitness* yang dihasilkan. Nilai *fitness* terbaik yang dihasilkan pada proses *training* sangat rendah, hanya -138.71074. Hal ini menjadikan pengujian ini sebagai pengujian dengan nilai *fitness* terendah kedua sejauh ini yang menggunakan *fitness* *function 2*, hanya Tahap 1 Pengujian 2 yang memiliki nilai *fitness* lebih rendah. Belum lagi, kompleksitas *network* juga menjadi tinggi. Grafik nilai *fitness* dan rata-rata nilai *fitness* dalam populasi pada pengujian ini dapat dilihat pada Gambar 5.20

A close up of text on a white background

Description automatically generated

Gambar 5.20 Grafik nilai fitness terbaik dan nilai fitness rata-rata dalam populasi selama proses training pada Tahap 3 Pengujian 1

Efek negatif juga masih terasa pada saat proses testing. Network yang dihasilkan pada pengujian ini hanya mampu memprediksi hasil pertandingan dengan benar sebanyak 260 dari 380 pertandingan, dimana hanya 90 pertandingan yang skornya dapat ditebak dengan benar. Hasil yang kurang memuaskan pada pengujian ini dapat disebabkan oleh *encoding* yang tidak cocok, atau posisi pemain sulit diproses oleh network sehingga akurasi prediksi menjadi rendah. Rangkuman proses *training* dan *testing* dari pengujian ini dapat dilihat pada Tabel 5.7.

Tabel 5.7 Rangkuman proses training dan testing pada Tahap 3 Pengujian 1

|  |  |
| --- | --- |
| Generasi | 5000 |
| Rata-rata waktu per generasi | 26.809 detik |
| Total waktu | 2234.08333 menit |
| Jenis *Encoding* | Label *Encoding* |
| *Fitness Function* | *Fitness Function 2* (Segmen Program 4.5) |
| Rata-rata nilai *fitness* dalam populasi | -11009.71893 |
| Nilai *fitness* terbaik dalam populasi | -138.71074 |
| Kompleksitas *network* pada individu dengan nilai *fitness* terbaik | 71 *nodes* dan 137 koneksi aktif |
| Skor pertandingan benar | 90 (23.6%) |
| Hasil pertandingan benar | 260 (68.4%) |

#### Tahap 3 Pengujian 2

Karena hasil yang tidak memuaskan pada Tahap 3 Pengujian 2, pada pengujian ini jenis *encoding* akan diganti. Jenis *encoding* yang akan digunakan pada pengujian ini adalah *onehot encoding*. Terdapat 16 posisi berbeda yang dapat ditempati pemain, yang menjadikan jumlah *feature* pada pengujian ini menjadi sangat banyak. Total jumlah *feature* yang digunakan pada pengujian ini adalah 376, yang terdiri dari 22 *rating* pemain, 2 *rating* *team*, dan 22 posisi pemain yang setelah di*encode* menjadi 352 (22 \* 16).

A close up of text on a white background

Description automatically generated

Gambar 5.21 Konfigurasi yang digunakan pada Tahap 3 Pengujian 2

Konfigurasi yang digunakan juga mengalami penyesuaian karena jumlah *feature* yang terlalu besar.Seperti pop\_size yang mengatur besarnya jumlah populasi diturunkan menjadi 700 dari 1000, dan initial\_connection partial\_direct, yang mengatur koneksi awal pada masing-masing individu juga diturunkan menjadi 0.4 dari yang sebelumnya 0.7.

A screenshot of a cell phone

Description automatically generated

Gambar 5.22 Hasil training dari Tahap 3 Pengujian 2

Berbeda dari pengujian-pengujian sebelumnya, proses *training* pada pengujian ini dihentikan pada 1000 generasi. Pengujian dihentikan karena *progress* dari nilai *fitness* dan rata-rata nilai *fitness* dalam populasi sangat lambat, dan waktu yang dibutuhkan untuk menjalankan 1 generasi sangat lama. Dapat dilihat pada Gambar 5.22., nilai *fitness* terbaik yang dihasilkan sangat rendah, yaitu -3384.64161 dan waktu yang untuk menyelesaikan 1 generasi lebih dari 50 detik. Untuk menjalankan 1000 generasi saja, total waktu yang dibutuhkan adalah 926.53333 menit. Selain itu, kompleksitas *network*, terutama koneksi menjadi sangat banyak karena banyaknya *feature* yang digunakan. Grafik nilai *fitness* terbaik dan nilai *fitness* rata-rata dalam populasi dapat dilihat pada Gambar 5.23.

A close up of text on a white background

Description automatically generated

Gambar 5.23 Grafik nilai fitness terbaik dan nilai fitness rata-rata dalam populasi selama proses training pada Tahap 3 Pengujian 2

Selaras dengan proses training, tingkat akurasi yang dihasilkan proses testing juga tidak memuaskan. Network hanya mampu memprediksi hasil pertandingan dengan benar sebanyak 194 dari 380 pertandingan, dimana 47 diantarnya dapat diprediksi dengan skor yang benar. Rangkuman proses *training* dan *testing* pada pengujian ini dapat dilihat pada Tabel

Tabel 5.8 Rangkuman proses training dan testing pada Tahap 3 Pengujian 2

|  |  |
| --- | --- |
| Generasi | 1000 |
| Rata-rata waktu per generasi | 55.592 detik |
| Total waktu | 926.53333 menit |
| Jenis *Encoding* | *OneHot* *Encoding* |
| *Fitness Function* | *Fitness Function 2* (Segmen Program 4.5) |
| Rata-rata nilai *fitness* dalam populasi | -5281.77021 |
| Nilai *fitness* terbaik dalam populasi | -3384.64161 |
| Kompleksitas *network* pada individu dengan nilai *fitness* terbaik | 393 *nodes* dan 478 koneksi aktif |
| Skor pertandingan benar | 47 (12 %) |
| Hasil pertandingan benar | 194 (51 %) |

#### Tahap 3 Pengujian 3

Karena hasil dari Tahap 3 Pengujian 2 masih belum memuaskan juga, pada pengujian kali ini jenis *encoding* akan kembali diganti. Jenis *encoding* yang akan digunakan pada pengujian ini adalah *binary encoding*. *Encoding* jenis ini akan merubah merubah nilai posisi pemain menjadi sebuah nilai *binary*. Keuntungan *encoding* jenis ini dibanding *onehot encoding* adalah dimensi *output* yang dihasilkan lebih sedikit. Karena ada 16 posisi pemain, maka *output* dari *binary encoding* hanya memiliki 4 dimensi, dibandingkan 16 dimensi dari *output* yang dihasilkan *onehot encoding*. Total ada 112 *feature* yang digunakan pada pengujian ini, yang terdiri dari 22 *rating* pemain, 2 *rating* team, dan 88 nilai posisi pemain (22 \* 4).

A close up of a newspaper

Description automatically generated

Gambar 5.24 Konfigurasi yang digunakan pada Tahap 3 Pengujian 3

Penyesuaian konfigurasi juga kembali dilakukan pada penelitian ini. Selain num\_inputs yang berganti menjadi 112, pop\_size yang mengatur jumlah populasi juga kembali ditingkatkan menjadi 1000 karena jumlah *feature* tidak sebanyak Tahap 3 Pengujian 2. Kemampuan *network* untuk kehilangan koneksi dan *nodes* yang diatur oleh conn\_delete­\_prob dan nodes­\_delete\_prop juga dikembali diaktifkan, dan diatur ke 0.4.

A screenshot of a cell phone

Description automatically generated

Gambar 5.25 Hasil training dari Tahap 3 Pengujian 3

Sama seperti pengujian sebelumnya, proses training pengujian ini juga dihentikan pada 1000 generasi. Alasan dihentikannya proses *training* ini juga sama, yaitu waktu yang terlalu lama disertai *progress* dari nilai *fitness* sangat lambat. Waktu yang dibutuhkan untuk menyelesaikan 1 generasi pada proses *training* ini sekitar 29 detik. Nilai fitness yang dihasilkan proses *training* pada pengujian ini juga terbilang kecil, hanya -1560.5453 dengan rata-rata nilai *fitness* pada populasi sebesar -12364.50885. Grafik *fitness* terbaik selama pengujian dan rata-rata *fitness* dalam populasi dapat dilihat pada Gambar 5.26.

Hasil dari proses *testing* pada pengujian ini juga kurang memuaskan. *Network* hanya mampu menebak hasil pertandingan dengan benar sebanyak 223 dari 380 pertandingan, dimana 46 pertandingan diantaranya dapat ditebak dengan skor yang benar. Tingkat akurasi yang dihasilkan pengujian ini hanya sedikit lebih baik daripada Tahap 3 Pengujian 2 dan masih tertinggal cukup jauh daripada pengujian dengan tingkat akurasi terbaik yang pernah dilakukan, yaitu pada Tahap 2 Pengujian 2. Rangkuman proses *training* dan *testing* pada pengujian ini dapat dilihat pada Tabel 5.9.

A close up of a piece of paper

Description automatically generated

Gambar 5.26 Grafik nilai fitness terbaik dan nilai fitness rata-rata dalam populasi selama proses training pada Tahap 3 Pengujian 3

Tabel 5.9 Rangkuman proses training dan testing pada Tahap 3 Pengujian 3

|  |  |
| --- | --- |
| Generasi | 1000 |
| Rata-rata waktu per generasi | 28.952 detik |
| Total waktu | 482.53333 |
| Jenis *Encoding* | *Binary* *Encoding* |
| *Fitness Function* | *Fitness Function 2* (Segmen Program 4.5) |
| Rata-rata nilai *fitness* dalam populasi | -12364.50885 |
| Nilai *fitness* terbaik dalam populasi | -1560.54523 |
| Kompleksitas *network* pada individu dengan nilai *fitness* terbaik | 128 *nodes* dan 114 koneksi aktif |
| Skor pertandingan benar | 46 (12 %) |
| Hasil pertandingan benar | 1. (58 %) |

#### Tahap 3 Pengujian 4

Berbeda dengan pengujian lainnya yang dilakukan pada Tahap 3.,posisi setiap pemain pada pengujain ini tidak di*encode* satu persatu. Jumlah dari masing-masing *defender*, *midfielder*, dan *striker*lah yang menjadi parameter tambahan. Dengan begitu, jumlah *features* yang digunakan pada pengujian ini hanya 30, yang terdiri dari 11 *rating* pemain, 1 *ratimg* *team*, 1 jumlah *defender*, 1 jumlah *midfielder,* 1 jumlah *striker* dari masing masing *team*.

*A close up of text on a white background

Description automatically generated*

Gambar 5.27 Konfigurasi yang digunakan pada Tahap 3 Pengujian 4

Konfigurasi yang digunakan pada pengujian ini masih sama dengan konfigurasi yang digunakan pada Tahap 3 Pengujian 3., hanya saja jumlah input berkurang menjadi 30, yang menjadikan pengujian ini pengujian dengan jumlah *features* paling sedikit pada Tahap 3. Konfigurasi dapat dilihat pada Gambar 5.27.

A screenshot of a cell phone

Description automatically generated

Gambar 5.28 Hasil training dari Tahap 3 Pengujian 4

Ternyata, penggunaan *encoding* posisi pemain secara keseluruhan membawa hasil positif pada proses *training*. Selain lebih singkatnya waktu yang dibutuhkan untuk menyelesaikan satu generasi, nilai fitness terbaik yang dihasilkan, 27.67263, juga melebihi seluruh nilai fitness terbaik yang dihasilan oleh pengujian pada Tahap 3. Proses training ini dijalankan sebanyak 5000 generasi dengan dengan rata-rata waktu yang dibutuhkan sebanyak 25 detik pergenerasi. Grafik *fitness* terbaik selama pengujian dan rata-rata *fitness* dalam populasi dapat dilihat pada Gambar 5.29.

Hasil positif juga berlanjut pada proses *testing*. Akurasi yang dihasilkan pengujian ini juga melampaui semua akurasi yang dihasilkan oleh pengujian lain pada tahap ini, baik dalam akurasi hasil maupun akurasi skor pertandingan. *Network* yang dihasilkan oleh pengujian ini mampu menebak 299 dari 380 pertandingan dengan benar, dimana 131 diantaranya dapat ditebak dengan skor yang tepat.

Tabel 5.10 Rangkuman proses training dan testing pada Tahap 3 Pengujian 4

|  |  |
| --- | --- |
| Generasi | 5000 |
| Rata-rata waktu per generasi | 25.464 |
| Total waktu | 2122 |
| *Fitness Function* | *Fitness Function 2* (Segmen Program 4.5) |
| Rata-rata nilai *fitness* dalam populasi | -9819.25758 |
| Nilai *fitness* terbaik dalam populasi | 27.67263 |
| Kompleksitas *network* pada individu dengan nilai *fitness* terbaik | 42 *nodes* dan 114 koneksi aktif |
| Skor pertandingan benar | 131 (34 %) |
| Hasil pertandingan benar | 299 (78.6 %) |

A screenshot of a cell phone

Description automatically generated

Gambar 5.29 Grafik nilai fitness terbaik dan nilai fitness rata-rata dalam populasi selama proses training pada Tahap 3 Pengujian 4

### Kesimpulan Pengujian NEAT

Selain *feature*, *fitness* *function* juga memiliki peranan yang sangat penting pada NEAT. Dapat dilihat pada Tahap 1 Pengujian 1 dan Tahap 1 Pengujian 2, terjadi peningkatan akurasi yang signifikan. Pengujian dilakukan dengan *feature* dan konfigurasi yang sama, namun *fitness* *function* yang digunakan berbeda. Tingkat akurasi hasil pertandingan dan akurasi skor yang dihasilkan Tahap 1 Pengujian 1, yaitu 28% dan 8% meningkat menjadi 71% dan 22% pada Tahap 1 Pengujian 2.

Selain *fitness function* yang digunakan, konfigurasi juga memiliki peranan penting ketika proses NEAT berjalan. Pentingnya konfigurasi dapat dilihat dari Tahap 1 Pengujian 2 dan Tahap 1 Pengujian 3. Dengan menggunakan *feature* yang sama, yaitu *rating* pemain, terjadi peningkatan akurasi dari 71% untuk akurasi prediksi hasil pertandingan dan 22% untuk akurasi prediksi skor yang dihasilkan oleh Tahap 1 Pengujian 2, menjadi 78% dan 32% yang dihasilkan oleh Tahap 1 Pengujian 3.

Penambahan 2 *feature* baru, yaitu *rating* *team* juga ikut meningkatkan tingkat akurasi dari sebuah *network*, walaupun tidak signifikan. Dengan menggukana konfigurasi dan *fitness* *function* yang sama, akurasi prediksi hasil pertandingan dan akurasi prediksi skor meningkat dari 78% dan 32% yang dihasilkan oleh Tahap 1 Pengujian 3, menjadi 79% dan 35% pada Tahap 2 dan Pengujian 1.

Tingkat akurasi terbaik pada pengujian NEAT dihasilkan pada Tahap 2 Pengujian 2. Dengan sedikit melakukan perubahan konfigurasi yang digunakan pada Tahap 2 Pengujian 1, Tahap 2 Pengujian 2 mampu menghasilkan akurasi yang lebih baik, yaitu sebesar 81% untuk akurasi prediksi hasil pertandingan, dan 42% untuk akurasi prediksi skor.

Pada Tahap 3, posisi pemain ikut menjadi *feature* pada proses NEAT. Dari 3 pengujian yang mengaplikasikan *encoding* kepada posisi masing-masing pemain dan 1 pengujian yang menggunakan total dari posisi pemain pada setiap sektor, hasil terbaik diperoleh pada Tahap 3 Pengujian 4 yang menggunakan total posisi pemain pada tiap sektor. Pengujian ini menghasilkan tingkat akurasi sebesar 78% untuk akurasi hasil pertandingan dan 34% untuk akurasi skor.

Selain tingkat akurasi yang dihasilkan tidak sebaik Tahap 2, waktu yang dibutuhkan untuk menyelesaikan proses *training* juga jauh lebih lama. Pada Tahap 2, waktu yang dibutuhkan untuk menyelesaikan 1 generasi dalam proses *training* tidak pernah melebihi 20 detik. Sedangkan pada Tahap 3, waktu tercepat yang dibutuhkan untuk menyelesaikan 1 generasi adalah sekitar 25 detik.

Berdasarkan beberapa pengujian yang dilakukan dengan berbagai *features*, *fitness function*, dan konfigurasi yang berbeda, dapat disimpulkan bahwa pengujian yang menghasilkan tingkat akurasi terbaik adalah Tahap 2 Pengujian 3 dengan *features* berupa *rating* pemain dan *rating team*. Rangkuman pengujian dapat dilihat pada Tabel 5.11.

Tabel 5.11 Rangkuman seluruh pengujian NEAT

|  |  |  |
| --- | --- | --- |
| Pengujian | Akurasi Hasil Pertandingan | Akurasi Skor Pertandingan |
| Tahap 1 Pengujian 1 | 107 / 380 (28%) | 32 / 380 (8%) |
| Tahap 1 Pengujian 2 | 273 / 380 (71%) | 85 / 380 (22%) |
| Tahap 1 Pengujian 3 | 300 / 380 (78%) | 125 / 300 (32%) |
| Tahap 2 Pengujian 1 | 302 / 380 (79%) | 136 / 380 (35%) |
| Tahap 2 Pengujian 2 | 305 / 380 (80%) | 136 / 380 (35%) |
| Tahap 3 Pengujian 1 | 260 / 380 (68%) | 90 / 380 (23%) |
| Tahap 3 Pengujian 2 | 194 / 380 (51%) | 47 / 380 (12%) |
| Tahap 3 Pengujian 3 | 223 / 380 (52 %) | 46 / 380 (12%) |
| Tahap 3 Pengujian 4 | 299 / 380 (78.6 %) | 131 / 380 (34 %) |

## Pengujian *Backpropagation*

Setelah proses NEAT selesai dijalankan, proses selanjutnya adalah *backpropagation*. Proses *backpropagation* bertujuan untuk mengoptimasi *weight* dari *network* yang dihasilkan oleh NEAT. Hasil dari proses *backpropagation* dapat dilihat pada Tabel 5.12.

Tabel 5.12 Hasil pengujian backpropagation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Pengujian | *Error* Terakhir | *Epoch* | Akurasi Hasil Pertandingan | Akurasi Skor Pertandingan | Total Waktu |
| Tahap 1 Pengujian 1 | 759.55 | 100.000 | 291 / 380 (76%) | 94 / 380 (24%) | 264 detik |
| Tahap 1 Pengujian 2 | 460.110 | 100.000 | 289 / 380 (76%) | 134 / 380 (35%) | 234 detik |
| Tahap 1 Pengujian 3 | 433.55 | 44.000 | 301 / 380 (79%) | 155 / 380 (40 %) | 181 detik |
| Tahap 2 Pengujian 1 | 351.01 | 100.000 | 303 / 380 (79 %) | 167 / 380 (44 %) | 355 detik |
| Tahap 2 Pengujian 2 | 395.55 | 100.000 | 308 / 380 (81%) | 161 / 380 (42%) | 482 detik |
| Tahap 3 Pengujian 1 | 386.29 | 100.000 | 299 / 380 (78%) | 153 / 380 (40%) | 789 detik |
| Tahap 3 Pengujian 2 | 492.47 | 100.000 | 306 / 380 (80%) | 122 / 380 (32%) | 2269 detik |
| Tahap 3 Pengujian 3 | 469.00 | 100.000 | 304 / 380 (80%) | 162 / 380 (42%) | 1199 detik |
| Tahap 3 Pengujian 4 | 326.57 | 100.000 | 310 / 380 (81.5%) | 185 / 380 (48%) | 656  detik |

Semua pengujian *backpropagation* dijalankan dengan *learning rate* sebesar 0.0001 dan selama 100.000 *epoch*, kecuali untuk Tahap 1 Pengujian 3 karena terjadinya *overfitting* pada *epoch* 44.000 – 45.000 yang menyebabkan proses *backpropagation* terpaksa dihentikan.

Berdasarkan pengujian-pengujian yang telah dilakukan, proses *backpropagation* terbukti mampu meningkatkan tingkat akurasi pada *network* yang dihasilkan oleh NEAT pada penelitian ini dengan rata-rata peningkatan akurasi sebesar 15,25% pada akurasi hasil pertandingan, dan 12,625% pada akurasi skor pertandingan.

Peningkatan paling signifikan pada *network* setelah dilakukannya *backpropagation* terjadi pada Tahap 1 Pengujian 1. *Network* yang dihasilkan oleh Tahap 1 Pengujian 1 pada awalnya hanya memiliki akurasi sebesar 28% dan 8%. Namun setelah proses *backpropagation* diaplikasikan pada pengujian tersebut, tingkat akurasi mengalami kenaikkan menjadi 76% dan 24%. Hal ini membuktikan bahwa walaupun tingkat akurasi yang dihasilkan oleh *network* dari proses NEAT rendah, struktur yang dimiliki *network* itu sudah cukup baik.

Perbandingan tingkat akurasi sebelum dan sesudah dilakukannya *backpropagation* dapat dilihat pada Tabel 5.13.

Tabel 5.13 Perbandingan akurasi dari network sebelum dan sesudah backpropagation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pengujian | Sebelum  *Backpropagation* | | Sesudah  *Backpropagation* | |
| Akurasi Hasil | Akurasi Skor | Akurasi  Hasil | Akurasi  Skor |
| Tahap 1 Pengujian 1 | 107 / 380 (28%) | 32 / 380 (8%) | 291 / 380 (76%) | 94 / 380 (24%) |
| Tahap 1 Pengujian 2 | 273 / 380 (71%) | 85 / 380 (22%) | 289 / 380 (76%) | 134 / 380 (35%) |
| Tahap 1 Pengujian 3 | 300 / 380 (78%) | 125 / 300 (32%) | 301 / 380 (79%) | 155 / 380  (40 %) |
| Tahap 2 Pengujian 1 | 302 / 380 (79%) | 136 / 380 (35%) | 303 / 380  (79 %) | 167 / 380  (44 %) |
| Tahap 2 Pengujian 2 | 305 / 380 (80%) | 136 / 380 (35%) | 308 / 380 (81%) | 161 / 380 (42%) |
| Tahap 3 Pengujian 1 | 260 / 380 (68%) | 90 / 380 (23%) | 299 / 380 (78%) | 153 / 380 (40%) |
| Tahap 3 Pengujian 2 | 194 / 380 (51%) | 47 / 380 (12%) | 306 / 380 (80%) | 122 / 380 (32%) |
| Tahap 3 Pengujian 3 | 223 / 380  (52 %) | 46 / 380 (12%) | 304 / 380 (80%) | 162 / 360 (42%) |
| Tahap 3 Pengujian 4 | 299 / 380 (78.6 %) | 131 / 380 (34 %) | 310 / 380 (81.5%) | 185 / 380 (48%) |

Setelah *backpropagation* selesai dijalankan, pengujian yang menghasilkan *network* dengan akurasi terbaik berubah. Tahap 3 Pengujian 4 mengalahkan Tahap 2 Pengujian 2 yang sebelumnya menjadi pengujian dengan akurasi terbaik.

Akurasi yang dihasilkan oleh Tahap 3 Pengujian 4 meningkat dari yang sebelummya 78% untuk akurasi hasil pertandingan dan 34% akurasi skor pertandingan, menjadi 81% dan 48%. Akurasi yang dihasilkan oleh pengujian ini mengalahkan semua akurasi dari semua pengujian yang telah dilakukan.

## Pengujian *Real Life*

Pada data pertandingan *real*, tidak diketahui berapa *rating* pemain yang akan bermain pada pertandingan tersebut. Untuk memprediksi *rating* pemain yang akan bermain, akan dicoba beberapa cara, yaitu rata-rata 5 pertandingan terakhir, rata-rata 10 pertandingan terakhir.

Untuk rata-rata 5 dan 10 pertandingan terakhir, akan diambil *rating* dari pemain dan *team* pada 5 dan 10 pertandingan terakhir sebelum pertandingan akan dicoba untuk diprediksi. Sebagai contoh, pada pertandingan ke 20, akan dihitung rata-rata pada pertandingan ke-14 sampai ke-19 atau pertandingan ke-9 sampai pertandingan ke-19. Jika data yang tersedia kurang dari 5 atau 10, akan dihitung rata-rata dari jumlah pertandingan yang tersedia dalam rentang 5 atau 10 pertandingan terakhir. Jika seorang pemain tidak pernah bermain pada 5 atau 10 pertandingan terakhir, akan diberi *rating* 0.55.

Selain itu, akan diuji juga rata-rata 5 dan 10 pertandingan terakhir yang sudah dinormalisasi. Normalisasi dilakukan dengan cara mengalikan *rating* pemain dengan *rating* *team* lawan, lalu dibagi dengan *rating* *team* dari pemain itu sendiri.

Agar tersedianya cukup data, pengujian *real life* dimulai dengan pertandingan dari pertandingan ke-10 untuk rata-rata 5 pertandingan terakhir, dan pertandingan ke-15 untuk rata-rata 10 pertandingan terakhir.

Pengujian *real life* dilakukan dengan *network* hasil pengujain *backpropagation* yang memiliki tingkat akurasi tertinggi, yaitu *network* dari Tahap 3 Pengujian 4.

Tabel 5.14 Hasil pengujian real life

|  |  |  |
| --- | --- | --- |
| Metode Prediksi Rating | Akurasi Hasil Pertandingan | Akurasi Skor Pertandingan |
| Rata-rata 5 pertandingan terakhir | 124 / 280  (44%) | 28 / 280  (10%) |
| Rata-rata 5 pertandingan terakhir dengan normalisasi | 78 / 280  (27.8%) | 25 / 280  (8.9%) |
| Rata-rata 10 pertandingan terakhir | 92 / 230  (40%) | 21 / 230  (9%) |
| Rata-rata 10 pertandingan terakhir dengan normalisasi | 63 / 230  (26.5%) | 20 / 230  (8.6%) |

Seperti yang dapat dilihat pada Tabel 5.14., hasil pengujian *real life* sangat memiliki akurasi yang sangat rendah jika dibandingkan dengan hasil pengujian *backpropagation* maupun hasil pengujian NEAT yang menggunakan *rating* yang sebenarnya. Kedua metode yang digunakan untuk melakukan pengujian ini memperoleh tingkat akurasi yang hampir sama, yaitu 44% dan 10% untul rata-rata 5 pertandingan terkahir, 27.8% dan 8.9% jika dilakukan dengan normalisasi dan 40% dan 9% untuk rata-rata 10 pertandingan terakhir, 26.5% dan 8.6% jika dilakukan dengan normalisasi. Hal ini disebabkan oleh *rating* pemain yang tidak konsisten pada setiap pertandingan sehingga sulit untuk diprediksi menggunakan rata-rata.

## Kesimpulan Pengujian

Setelah semua pengujian selesai dijalankan, dapat disimpulkan bahwa *features* terbaik pada pengujian NEAT adalah *rating* pemain disertai dengan *rating team* yang terjadi pada Tahap 2 Pengujian 2 dengan tingkat akurasi sebesar 80% untuk akurasi hasil pertandingan, dan 35% untuk akurasi skor. Penggunaan rating pemain dan *rating* team sedikit meningkatkan tingkat akurasi jika *feature* yang digunakan hanya *rating* pemain.

Penambahan posisi pemain sebagai *features* pada NEAT justru menimbulkan efek negatif. Selain tingkat akurasi yang lebih rendah jika dibandingkan dengan penggunaan *features* lainnya, waktu *training* yang dibutuhkan juga jauh lebih lama.

Dalam beberapa kasus, struktur yang dihasilkan oleh NEAT sudah cukup baik, tetapi *weight* yang tidak optimal menyebabkan tingkat akurasi pada pengujian menjadi rendah. Hal ini dibuktikan dengan peningkatan akurasi yang signifikan pada *network* yang memiliki tingkat akurasi rendah ketika proses *backpropagation* dijalankan.

*Backpropagation* terbukti mampu meningkatkan tingkat akurasi pada semua *network* yang dihasilkan oleh NEAT. Pada *network* yang sudah memiliki tingkat akurasi tinggi, memang optimasi *backpropagation* tidak terlalu signifikan. Tetapi pada *network* yang memiliki akurasi rendah, *backpropagation* mampu mengoptimasi *weight* sehingga terjadi peningkatan akurasi yang signifikan. Peningkatan akurasi terbesar terjadi pada Tahap 1 Pengujian 1, dimana akurasi hasil pertandingan meningkat dari 28% menjadi 76%.

Setelah proses *backpropagation* selesai, Tahap 3 Pengujian 4 yang menambahkan total posisi tiap pemain pada berbagai sektor keluar sebagai pengujian yang menghasilkan tingkat akurasi terbaik, mengalahkan Tahap 2 Pengujian 2 yang sebelumnya memiliki tingkat akurasi terbaik. Hal ini kembali membuktikan bahwa adanya kemungkinan *weight* yang dihasilkan NEAT tidak optimal.

Sayangnya, akurasi yang didapat pada pengujian *real life* tidak sebaik akurasi yang dihasilkan pada proses *testing*. Karena pada pengujian *real life* *rating* pemain maupun *rating* *team* tidak diketahui, sehingga harus diprediksi terlebih dulu. Penggunaan metode rata-rata untuk menentukan *rating* pemain dan *rating* team gagal menghasilkan akurasi yang baik. Inkonsistensi dari pemain pada satu pertandingan ke pertandingan lainnya diyakini menjadi penyebab rendahnya akurasi yang dihasilkan.